Y4. Sci 2: 99-59

Science Policy Study—Hearings Volume 13
BRITISH SCIENCE EVALUATION METHODS

ORIGINAL

COMPLETED

H118 420

HEARING

BEFORE THE

TASK FORCE ON SCIENCE POLICY

OF THE

COMMITTEE ON
SCIENCE AND TECHNOLOGY
HOUSE OF REPRESENTATIVES

NINETY-NINTH CONGRESS

FIRST SESSION

OCTOBER 30, 1985

[No. 59]

Printed for the use of the
Committee on Science and Technology

Science Policy Study—Hearings Volume 13
BRITISH SCIENCE EVALUATION METHODS

HEARING

BEFORE THE

TASK FORCE ON SCIENCE POLICY

OF THE

COMMITTEE ON
SCIENCE AND TECHNOLOGY
HOUSE OF REPRESENTATIVES

NINETY-NINTH CONGRESS

FIRST SESSION

OCTOBER 30, 1985

[No. 59]

Printed for the use of the
Committee on Science and Technology

U.S. GOVERNMENT PRINTING OFFICE
56-399 O
WASHINGTON : 1986

I

COMMITTEE ON SCIENCE AND TECHNOLOGY

DON FUQUA, Florida, *Chairman*

ROBERT A. ROE, New Jersey
GEORGE E. BROWN, Jr., California
JAMES H. SCHEUER, New York
MARILYN LLOYD, Tennessee
TIMOTHY E. WIRTH, Colorado
DOUG WALGREN, Pennsylvania
DAN GLICKMAN, Kansas
ROBERT A. YOUNG, Missouri
HAROLD L. VOLKMER, Missouri
BILL NELSON, Florida
STAN LUNDINE, New York
RALPH M. HALL, Texas
DAVE McCURDY, Oklahoma
NORMAN Y. MINETA, California
MICHAEL A. ANDREWS, Texas
BUDDY MacKAY, Florida**
TIM VALENTINE, North Carolina
HARRY M. REID, Nevada
ROBERT G. TORRICELLI, New Jersey
FREDERICK C. BOUCHER, Virginia
TERRY BRUCE, Illinois
RICHARD H. STALLINGS, Idaho
BART GORDON, Tennessee
JAMES A. TRAFICANT, Jr., Ohio

MANUEL LUJAN, Jr., New Mexico*
ROBERT S. WALKER, Pennsylvania
F. JAMES SENSENBRENNER, Jr., Wisconsin
CLAUDINE SCHNEIDER, Rhode Island
SHERWOOD L. BOEHLERT, New York
TOM LEWIS, Florida
DON RITTER, Pennsylvania
SID W. MORRISON, Washington
RON PACKARD, California
JAN MEYERS, Kansas
ROBERT C. SMITH, New Hampshire
PAUL B. HENRY, Michigan
HARRIS W. FAWELL, Illinois
WILLIAM W. COBEY, Jr., North Carolina
JOE BARTON, Texas
D. FRENCH SLAUGHTER, Jr., Virginia
DAVID S. MONSON, Utah

Harold P. Hanson, *Executive Director*
Robert C. Ketcham, *General Counsel*
Regina A. Davis, *Chief Clerk*
Joyce Gross Freiwald, *Republican Staff Director*

Science Policy Task Force

DON FUQUA, Florida, *Chairman*

GEORGE E. BROWN, Jr., California
TIMOTHY E. WIRTH, Colorado
DOUG WALGREN, Pennsylvania
HAROLD L. VOLKMER, Missouri
STAN LUNDINE, New York
NORMAN Y. MINETA, California
HARRY M. REID, Nevada
FREDERICK C. BOUCHER, Virginia
RICHARD H. STALLINGS, Idaho

MANUEL LUJAN, Jr., New Mexico*
TOM LEWIS, Florida
DON RITTER, Pennsylvania
SID W. MORRISON, Washington
RON PACKARD, California
JAN MEYERS, Kansas
HARRIS W. FAWELL, Illinois
D. FRENCH SLAUGHTER, Jr., Virginia

John D. Holmfeld, *Study Director*
R. Thomas Weimer, *Republican Staff Member*

---

*Ranking Republican Member.
**Serving on Committee on the Budget for 99th Congress.

CONTENTS

WITNESSES

October 30, 1985:

Page

Professor Benjamin R. Martin, science policy research unit, University of Sussex, Brighton, United Kingdom 2
- Prepared testimony 18
- Discussion 40

Appendix 1: Selected papers by Martin and Irvine:

Benjamin R. Martin and John Irvine, "Internal Criteria for Scientific Choice: An Evaluation of Research in High-Energy Physics Using Electron Accelerators," *Minerva*, vol. XIX, No. 3, autumn 1981, pp. 408-432 51

John Irvine and Ben R. Martin, "Assessing Basic Research: The Case of the Isaac Newton Telescope," *Social Studies of Science*, vol. 13, 1983, pp. 49-86 77

Ben R. Martin and John Irvine, "Assessing basic research: Some partial indicators of scientific progress in radio astronomy," *Research Policy*, vol. 12, 1983, pp. 61-90 115

John Irvine and Ben R. Martin, "What Direction for Basic Scientific Research?", in *Science and Technology Policy in the 1980's and Beyond*, edited by Michael Gibbons, Philip Gummett, and Bhalchandra Udgaokar (Longman, London, 1984), pp. 67-98 145

Ben R. Martin and John Irvine, "CERN: Past performance and future prospects. I. CERN's position in world high-energy physics", *Research Policy*, vol. 13, 1984, pp. 183-210 177

John Irvine and Ben R. Martin, "CERN: Past performance and future prospects. II. The scientific performance of CERN accelerators," *Research Policy*, vol. 13, 1984, pp. 247-284 205

Ben R. Martin and John Irvine, "CERN: Past performance and future prospects. III. CERN and the future of world high-energy physics", *Research Policy*, vol. 13, 1984, pp. 311-342 243

John Irvine and Ben R. Martin, "Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators," *Social Studies of Science*, vol. 15, 1985, pp. 293-341 275

John Irvine, Ben Martin, Tim Peacock and Roy Turner, "Charting the decline in British science," *Nature*, vol. 316, August 15, 1985, pp. 587-590 324

Michiel Schwartz, John Irvine, Ben Martin, Keith Pavitt, and Roy Rothwell, "The assessment of Government support for industrial research: lessons from a study of Norway," *R&D Management*, vol. 12, No. 4, 1982, pp. 155-167 333

Appendix 2: Critiques of Irvine and Martin's methodology and a reply by Martin and Irvine:

Robert Walgate, "Poor marks for enterprise," *Nature*, vol. 311, September 6, 1984, p. 4 346

V. F. Weisskopf, "Poor Marks for Enterprise?", submitted to *Nature* 348

John Krige and Dominique Pestre, "A Critique of Irvine and Martin's Methodology for Evaluating Big Science," *Social Studies of Science*, vol. 15, 1985, pp. 525-539 352

H. F. Moed and A. F. J. van Raan, "Critical Remarks on Irvine and Martin's Methodology for Evaluating Scientific Performance," *Social Studies of Science*, vol. 15, 1985, pp. 539-547 366

Robert Budd, "The Case of the Disappearing Caveat: A Critique of Irvine and Martin's Methodology," *Social Studies of Science*, vol. 15, 1985, pp. 548-553 375

BEST COPY AVAILABLE

Page

Appendix 2—Continued
  H. M. Collins, "The Possibilities of Science Policy," *Social Studies of Science*, vol. 15, 1985, pp. 554-558 381
  Ben R. Martin and John Irvine, "Evaluating the Evaluators: A Reply to Our Critics," *Social Studies of Science*, vol. 15, 1985, pp. 558-575 385

BRITISH SCIENCE EVALUATION METHODS

WEDNESDAY, OCTOBER 30, 1985

House of Representatives,
Committee on Science and Technology,
Task Force on Science Policy,
Washington, DC.

The task force met, pursuant to notice, at 8:35 a.m., in room 2168, Rayburn House Office Building, Hon. Don Fuqua (chairman of the task force) presiding.

Mr. Fuqua. We are delighted to welcome Prof. Benjamin Martin from the University of Sussex in England to the Science Policy Task Force. Professor Martin is a member of the Science Policy Research Unit at his university, and we are taking advantage of the fact that he is spending a few days in Washington to bring him before our task force.

I want to express to Professor Martin our thanks for his willingness to appear before us and to prepare a written statement on very short notice. We regret that his colleague and close collaborator, Prof. John Irvine, also of the Science Policy Research Unit at Sussex, is unable to be here today as well. Professor Irvine is currently in Japan as part of an important research project by the Martin-Irvine team.

In developing the agenda for our science policy study, we noted by way of introduction that in many areas of science policy we would like to see an increase in the use of statistical data and quantitative analysis. Such statistics have in the past been limited to information about findings and manpower—or funds and manpower, and the analysis of this information has been at quite an unsophisticated level.

Instead we have, on many of the more difficult questions of science and policy, relied on the experience, judgment, and wisdom of the best people we could find in the scientific community. Often the evidence which we have had before us has been very anecdotal in nature, and as a former colleague of ours, Congressman Ray Thornton, the president of the University of Arkansas, once noted about anecdotal evidence—he said: "No one doubts its veracity; what we need to determine is its representativeness."

Wisdom and anecdotal evidence will no doubt continue to play important roles in science policymaking. But it now appears that the first steps toward a broader use of quantitative approaches are being taken. In this area our British friends, both at the University of Sussex and in the British Research Councils which are funding and using some of this work, appear to be well in the lead. It is to

BEST COPY AVAILABLE

tell us about the important work that Professor Martin has agreed to appear.

We recognize that the methods developed in England have not yet been perfected. We are also aware that in some cases the results produced have been controversial both in England and more generally in Europe. Perhaps that makes us even more willing to welcome you, Professor Martin, and we look forward to your statement and the discussion to follow.

Mr. Lujan?

Mr. **Lujan**. I have no opening statement. I wanted to welcome Professor Martin. I look forward to hearing from him.

STATEMENT OF BEN R. MARTIN, PROFESSOR, SCIENCE POLICY RESEARCH UNIT, UNIVERSITY OF SUSSEX, ENGLAND

Professor **Martin**. Mr. Chairman, let me begin by thanking you for the invitation to testify here today. What I have to say relates to one issue in particular raised in the agenda of the Task Force on Science Policy; namely, that in past and current science policymaking there is very limited use of quantitative information.

There is more than a little irony in the fact that while science itself has been affected so immeasurably from the application of systematic, rigorous, and generally quantitative techniques, science policy is one of the last areas of public policy where such techniques have been applied. That there are now available quantitative techniques capable of yielding information relevant to science policymaking is what I hope to demonstrate.

Since 1978 a small team at the Science Policy Research Unit, SPRU, has carried out a range of quantitative science policy research studies. Much of our effort has been concentrated on developing systematic methods for evaluating the research performance of laboratories, facilities, and groups.

Initially this work focused on basic science, in particular on the large central facilities used to carry out research in big sciences such as high-energy physics—the results are written up in references contained in appendix 2—optical astronomy, and radio astronomy.

As I shall describe shortly, the overall approach adopted in such institutional evaluations involves the combined use of, one, a range of science output indicators based on an analysis of the published scientific literature to assess the productivity and impact of each research facility relative to similar facilities elsewhere; and, two, the results of structured interviews with a large sample of researchers in different countries who are asked to rank these facilities in terms of their relative contributions to scientific knowledge and to explain the factors determining differences in performance.

Over the last 4 years, however, our work at SPRU has diversified considerably. First, besides evaluating the outputs from research institutions, we have started to develop a program of work on the assessment of national scientific performance in different fields and subfields of science, especially those with strategic technological importance for the future. This work has mainly been commissioned—for example, by the British Advisory Board for the Research Councils and various learned societies—and covers such fields as protein crystallography and ocean currents as well as Britain's overall scientific performance relative to other countries. Recently published results on Britain's declining scientific performance attracted significant public interest in the UK earlier this year, notably during debates in the House of Lords and House of Commons.

Second, a number of studies commissioned by mainly foreign governmental agencies took us into the evaluation of applied R&D activities. These include projects carried out for a Norwegian Royal Commission on Industrial Research—we assessed the performance of research institutes and support mechanisms for electronics and mechanical engineering—and for the European Commission we helped in an evaluation of the European Community's Steel Research Program, as well as preparatory work for a current review of British support for engineering research by the Science and Engineering Research Council. In evaluating applied research, the approach normally involves using a somewhat different set of indicators based upon industrial and technological impacts, although scientific literature-based statistics were used in a recent study comparing international performance in the field of integrated optics.

Third, in line with the needs of policymakers to move beyond evaluations of past performance, we have begun to explore methods for systematically appraising the future prospects of major new facilities such as accelerators, work which I shall discuss briefly later.

More importantly, the question of how one selects long-term research priorities in the most effective manner was taken further in a 1984 study for the British Advisory Council on Applied Research and Development which has become interested in the question of strategic monitoring and forecasting of research and development. This study, the results of which we published as a book, *Foresight in Science: Picking the Winners*, involved a survey of state-of-the-art methods used by governments, research funding agencies, and high technology firms to formulate their priorities for longer term strategic research.

Finally, over the last year we have initiated a program of work on the inputs to research. The need for reliable input data—on funding, numbers of researchers, and so on—in order to understand and interpret research output data, especially those on national scientific performance in particular fields, has become ever more apparent over time. We are currently engaged in a major study to compare government funding of academic and academically related research across six countries—the United States, Japan, West Germany, France, the United Kingdom, and the Netherlands.

The aim is to produce not just totals for each country, but also a breakdown by 9 fields and 40 subfields. This is a very topical subject at present since senior policymakers are increasingly interested in knowing how their spending compares with that in other countries, and the international statistics available at present are somewhat limited technically, and in their level of disaggregation. This study is largely being financed by the British Advisory Board for the Research Councils who wish to know whether Britain's apparent decline in scientific performance is related to comparative levels of funding across countries. It will be completed early in 1986.

Now in order to illustrate the type of approach used in our science policy research studies, we thought it would be of greatest utility to the task force to focus in detail upon one of our recent evaluations of big science facilities, since the question of evaluating institutional productivity, while often politically contentious, is, as pointed out on page 14 of the task force agenda, a central question facing those concerned with science policy. Incidentally, for those of you interested in details of our other work, these can be found in the papers listed in appendix 2, and I've given copies of some of these to John Holmfeld.

First, however, let me review briefly some of the main reasons behind the need for better techniques for evaluating research performance. In Western countries, the main decisionmaking mechanism in basic science continues to be peer review; that is, relying on the opinions of scientists in the field concerned, relying on the informed prejudices of wise men, as it perhaps is best described.

However, peer review faces several problems of growing importance, especially in relation to big science. One stems from the trend away from the rapidly increasing national science budgets seen in the 1950's and 1960's toward approximately level funding, a trend which can only become more pronounced here in the United States as you grapple with the problem of the budget deficit.

Under the conditions of approximately zero real growth, decisions to reduce existing financial commitments often first have to be made in order to free funds to support promising new research areas and young scientists, something that you must do if science is to remain dynamic. This is not a task for which peer review is always very effective.

Second, as the costs of certain research have escalated and resources have become concentrated in fewer laboratories, so it has become harder to locate disinterested peers able to provide expert judgments on proposed new projects but whose own circumstances will be unaffected by the subsequent funding decision.

Third, because specialties generously funded in the past are normally well represented on decisionmaking bodies, peer review has exhibited a tendency to reproduce past priorities and to be less successful at identifying newly emerging—especially interdisciplinary—research areas.

Because of these problems, there is, we would argue, a need for more systematic output data to complement but not replace peer review. To illustrate how evaluations of research output might be carried out, I shall describe a study which compared the scientific performance of high-energy physics accelerators around the world, including those in U.S. laboratories.

How can one evaluate research performance in basic science? There are four main elements to our methodology. First, it is based on an input-output approach. It involves identifying and evaluating the various inputs—funds, researchers, et cetera—and outputs—for example, contributions to scientific knowledge, education, and technology—and then relating the two. For basic science, some simplification is possible since the primary output is contributions to scientific knowledge, and I shall, therefore, focus here on the evaluation of these scientific contributions.

Second, the approach here is institutionally focused—that is, the unit of analysis is the research laboratory, facility, or group—because major investment decisions in basic science normally center on institutions, and it is here that rigidities with peer review most often arise.

Third, because no absolute quantification of research performance is possible, the approach is comparative, with the proviso that one can only legitimately compare "like" with "like."

Fourth, the approach involves the combined use of several indicators, each reflecting a slightly different facet of research performance. For example, if one looks at the number of scientific papers published in international learned journals—and scientists place great stress on such papers as the primary means of making public their research results, so virtually every significant research finding is reported in a journal article—then the numbers of such papers give some indication of the scientific production of a research group, while numbers of papers per researcher or per dollar reveal something about its productivity.

Another important indicator is the average number of times each paper is referred to or cited by other scientists in subsequent papers. A typical paper will cite 10 or 20 previous papers which have had some impact on the work it reports. Thus, the important papers tend to be highly cited. For example, the most highly cited paper in experimental high-energy physics during the 1960's was also the only one which subsequently led to a Nobel Prize.

The average number of citations earned by each paper for a research group—and you can look these up in a computerized directory of citations published by the Institute for Scientific Information, Philadelphia—this gives an indication of the impact those publications have on the scientific community, while peer rankings for scientists asked to rank the relative contributions to science of different institutions provide evidence on the perceived significance of the results.

Last, data on the distribution of highly cited papers reveal which groups have been responsible for the few key discoveries in a specialty, while citation totals reflect the large number of incremental additions to knowledge.

It should be stressed that all these indicators are imperfect or partial. They reflect partly the relative magnitude of contributions to scientific knowledge and partly a variety of social, institutional, psychological, and other factors.

However, when applied to such research groups using similar research facilities, publishing in the same body of international literature, subject to comparable refereeing procedures, the indicators have been found in our various studies to yield broadly convergent results. Such results we regard as being reasonably reliable—certainly more reliable than those based on a single indicator like peer review.

Our study focused on the world's main proton accelerators, although output data relating to electron accelerators were also produced. From table 1—the tables are in appendix 1—you can see that in 1959 the CERN Proton Synchrotron, or PS, in the joint European laboratory at Geneva took over from the east European Dubna accelerator, and before that the Berkeley Bevatron in California, as the world's highest energy accelerator.

TABLE 1

The world's main proton accelerators (>5 GeV)

Accelerator	Began operating	Beam energy (GeV)
Berkeley Bevatron (U.S.)	1954	6
JINR Dubna (E.Europe)	1957	10
CERN PS (W.Europe)	1959	23
Brookhaven AGS (U.S.)	1960	33
ITEP Moscow (U.S.S.R.)	1961	7
Argonne ZGS (U.S.)	1963	12
Rutherford Nimrod (U.K.)	1963	7
Serpukhov (U.S.S.R.)	1967	76
CERN ISR (W.Europe)	1971	31
Fermilab (U.S.)	1972	400
CERN SPS (W.Europe)	1976	400
CERN $p\bar{p}$ (W.Europe)	1981	270

However, the very similar and slightly higher energy Brookhaven Alternating Gradient Synchrotron, or AGS, was completed a few months later in 1960. The lead passed to the Soviet Union in 1967 when the Serpukhov accelerator was completed, but returned to the United States in 1972 when the Fermilab accelerator began operating, 4 years ahead of the similar SPS accelerator at CERN.

Last, the CERN proton-antiproton, or $p\bar{p}$ collider started operation in 1981, and this remains the world's most powerful facility to date, although the larger Fermilab Tevatron Collider is very shortly to begin experimental work.

In analyzing the outputs from the accelerators, we divided the period from 1961 to 1982 into 4-year blocks. Thus, table 2, covering

BEST COPY AVAILABLE

1961 to 1964, shows that the Brookhaven AGS produced 4.5 percent of the world total of experimental high-energy physics papers in the first 2 years and 11 percent for 1963–64, somewhat behind its main rival, the CERN PS with 29.5 percent, and the older Berkeley Bevatron with 23 percent.

TABLE 2

Experimental high-energy physics, 1961-1964

	% of papers published in past two years[2]		% of citations to work of past four years[2]	Average citations per paper	Highly cited papers: number cited n times[1]			
	1962	1964	1964	1964	$n \geq 15$	$n \geq 30$	$n \geq 50$	$n \geq 100$
Bevatron	38.0%	23.0%	35.0%	3.4	33	5	2	0
Dubna	18.5%	12.0%	3.0%	0.5	0	0	0	0
CERN PS	10.0%	29.5%	14.5%	1.9	9	1	0	0
Brookhaven AGS	4.5%	11.0%	23.0%	8.0	24	5	1	1
Moscow ITEP	2.0%	3.0%	0.5%	0.5	0	0	0	0
Rest of world	27.0%	22.5%	24.0%	2.8	22	6	1	0
World total	375 100%	535 100%	2590 100%	2.8	88	17	4	1

1. Over the period 1961-80, an experimental paper required approximately 40 or more citations in any one year to be included in the top 1% most highly cited papers for the two decades, and 19 or more to be included in the top 5%. In this respect, $n \geq 15$ corresponds to the top 7.8%, $n \geq 30$ to the top 1.9%, $n \geq 50$ to the top 0.6%, and $n \geq 100$ to the top 0.14% most highly cited papers.

2. All figures in this and subsequent tables have been rounded to the nearest 0.5%.

However, the CERN papers reported data from relatively simple experiments, so their overall impact, as reflected in the number of times they were cited by other scientists, was less than that of the AGS and the Bevatron, the respective world citation shares being 14.5 percent, 23 percent, and 35 percent in 1964.

On average, each AGS paper was cited 8 times in 1964, a typical figure for a new accelerator, while CERN PS publications earned only 1.9 citations per paper. Moreover, in terms of highly cited papers, the Bevatron and AGS both seem to have been responsible for more discoveries than the PS. Indeed, the three major discoveries that could have been made on either the AGS or the PS—the identification of two types of neutrinos, the omega minus, and charge-parity violation—were all made on the U.S. accelerator.

Over the next 4 years, table 3 shows that the gap between the AGS and PS narrowed. The AGS continued to yield fewer papers—19.5 percent compared with 25.5 percent in 1968—but those papers were still more highly cited, averaging 4.6 citations per paper in 1968 compared with 3.1 for CERN publications, with the result that the total impact of the two machines seems to have been very similar, at least in terms of world citation share—the figures were 26 and 26.5 percent in 1968.

TABLE 3

Experimental high-energy physics, 1965-1968

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers: number cited n times			
	1966	1968	1966	1968	1966	1968	n≥15	n≥30	n≥50	n≥100
Bevatron	12.5%	11.5%	17.5%	10.0%	3.8	2.8	9	0	0	0
Dubna	6.5%	4.0%	2.5%	1.0%	1.1	0.9	0	0	0	0
CERN PS	32.5%	25.5%	28.5%	26.0%	3.5	3.1	46	1	0	0
Brookhaven AGS	19.5%	19.5%	28.0%	26.5%	6.8	4.6	42	12	1	0
Argonne ZGS	2.0%	6.5%	1.0%	5.0%	4.2	3.6	6	0	0	0
Rutherford Nimrod	2.5%	2.5%	2.5%	3.5%	6.0	4.5	9	0	0	0
Rest of world	24.0%	30.5%	20.0%	28.0%	3.3	3.4	38	5	0	0
World total	645 100%	845 100%	4500 100%	5080 100%	3.8	3.4	150	18	1	0

For papers cited 15 or more times in a year, the AGS had by then been overtaken by the PS, but the AGS continued to prove much more successful as regards the more important advances, with 12 papers cited 30 or more times compared with one from the PS.

For the period 1969-72, table 4 reveals that while the PS produced a steady 25 percent of world experimental publications, the AGS's share dropped to 14.5 percent in 1972, following difficulties associated with a major technical upgrade of the accelerator. However, the most highly cited papers at this time came from neither of these accelerators but from newer machines. The AGS and PS each managed only one paper cited 30 or more times, while Serpukhov in the Soviet Union, the CERN ISR, and the Stanford Linear Accelerator, SLAC, each yielded four.

TABLE 4

Experimental high-energy physics, 1969-1972

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers: number cited n times			
	1970	1972	1970	1972	1970	1972	n≥15	n≥30	n≥50	n≥100
Bevatron	9.0%	6.0%	10.0%	6.5%	2.9	2.5	3	1	0	0
CERN PS	25.0%	25.0%	22.5%	21.5%	2.6	2.3	19	1	0	0
Brookhaven AGS	19.0%	14.5%	23.0%	18.0%	3.5	2.9	22	1	0	0
Argonne ZGS	9.5%	9.5%	7.5%	9.0%	2.8	2.6	7	0	0	0
Rutherford Nimrod	2.5%	2.5%	2.5%	2.0%	3.2	1.9	0	0	0	0
Serpukhov	3.0%	6.0%	2.5%	5.0%	4.8	3.0	12	4	2	1
CERN ISR	-	2.0%	-	5.0%	-	12.9	13	4	2	0
SLAC	5.5%	8.0%	8.5%	12.5%	6.1	4.9	21	4	2	0
Rest of world	27.0%	26.0%	23.0%	19.5%	2.4	2.0	21	6	3	0
World total	950 100%	1020 100%	5270 100%	5310 100%	2.9	2.7	123	21	9	1

The next 4 years, 1973 to 1976, were among the most tumultuous ever in high-energy physics, ushering in the era of new physics, as it's called. CERN made a promising start in 1973 with the neutral currents discovery and several other major advances, but after this most of the important discoveries were made in the United States. As you can see from table 5, in terms of papers cited 15 times in a year, the PS and ISR CERN were well behind SLAC and the new Fermilab accelerator near Chicago, while no less than five out of eight crucial discoveries, cited 100 times in a year, came from the SPEAR collider at Stanford.

TABLE 5

Experimental high-energy physics, 1973-1976

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers : number cited n times			
	1974	1976	1974	1976	1974	1976	n≥15	n≥30	n≥50	n≥100
CERN PS	20.0%	23.5%	19.5%	14.5%	2.7	2.2	18	4	2	0
Brookhaven AGS	13.5%	8.0%	11.0%	10.0%	2.5	3.2	15	5	2	1
Argonne ZGS	9.0%	6.0%	7.0%	4.5%	2.3	2.1	5	0	0	0
Rutherford Nimrod	3.0%	1.5%	2.0%	0.5%	2.1	1.1	0	0	0	0
Serpukhov	9.5%	10.5%	7.0%	6.5%	2.8	2.1	7	0	0	0
CERN ISR	3.5%	5.0%	12.0%	9.0%	14.0	7.4	28	12	2	1
Fermilab	9.5%	15.0%	11.5%	25.0%	6.9	6.7	71	26	9	0
SLAC	7.0%	9.0%	10.0%	17.0%	4.2	7.1	37	14	8	5
Rest of world	26.0%	21.5%	20.0%	12.5%	2.4	1.8	16	6	3	1
World total	1150 100%	1180 100%	6780 100%	7740 100%	3.1	3.3	197	57	26	8

It may be of interest to note that the proposal to build this highly novel and successful accelerator at SLAC was turned down by the U.S. high-energy physics community on several occasions, illustrating the conservative tendencies implicit in peer review. As a result, Stanford had to resort to building the machine from its operating budget.

In addition, it was the Brookhaven AGS rather than the CERN PS which shared with Stanford the honor of making arguably the most important advance of the 1970's--the discovery of the J/psi particle which paved the way for the new physics. This yielded the most highly cited experimental paper of the 1970's, and, again, was the only discovery in the field to win the Nobel Prize.

In contrast, the following 4 years were a period of consolidation. According to the figures in table 6, the Fermilab accelerator seems to have contributed most during this time with about a quarter of total world citations and 40 papers cited 15 or more times. However, its performance did decline markedly toward the end of the period, partly as a result of funding problems, and partly because of the growing impact of its European rival, the new CERN SPS which, taking advantage of beams and detectors significantly better than those at Fermilab, achieved a particularly high rate of citations per paper, 12.7 in 1978 compared to 7.3 for Fermilab.

TABLE 6

Experimental high-energy physics, 1977-1980

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers: number cited n times			
	1978	1980	1978	1980	1978	1980	n≥15	n≥30	n≥50	n≥100
CERN PS	22.0%	11.5%	14.5%	12.5%	2.2	2.2	13	2	1	0
Brookhaven AGS	5.5%	5.5%	5.0%	3.0%	2.7	1.6	0	0	0	0
Serpukhov	12.0%	14.0	4.0%	5.0%	1.2	1.2	0	0	0	0
CERN ISR	4.5%	5.5%	7.0%	7.5%	5.4	4.4	11	2	0	0
Fermilab	16.5%	19.0%	32.0%	21.5%	7.3	3.6	40	10	5	1
CERN SPS	2.5%	8.5%	4.0%	8.5%	12.7	5.0	19	7	3	0
SLAC	9.5%	6.0%	15.0%	11.5%	5.7	4.4	26	6	1	1
DESY	4.0%	6.5%	5.5%	15.5%	5.7	8.8	36	16	4	0
Rest of world	23.5%	24.0%	13.0%	15.0%	2.0	1.9	19	5	0	0
World total	1115 100%	930 100%	8190 100%	6090 100%	3.5	3.0	164	48	14	2

As in previous periods, though, the crucial discoveries were made in the United States rather than at CERN, the two papers cited over 100 times during the 4 years reporting the discoveries of the upsilon at Fermilab and of parity violation at SLAC.

Finally, what has happened since 1980? Comparison of table 7 with earlier tables shows that Fermilab's world share of citations had fallen from a peak of 32 percent in 1978 to 16 percent in 1982, while the figure for SLAC had likewise dropped by nearly half between 1976, when it was 17 percent, and 1982, when it was 9 percent. Overall, the U.S. share of citations has decreased sharply from 59 percent of the world total in 1978 to 34 percent in 1982.

TABLE 7

Experimental high-energy physics, post-1980

	% of papers published		% of citations to work of past four years		Average citations per paper		Highly cited papers : number cited n times	
	1981	1982	1981	1982	1981	1982	n≥15	n≥30
CERN PS	9.5%	7.0%	8.5%	5.0%	1.6	1.4	0	0
Brookhaven AGS	4.5%	3.0%	3.0%	2.5%	1.7	1.4	0	0
Serpukhov	15.0%	11.5%	6.5%	6.0%	1.4	1.3	0	0
CERN ISR	6.0%	9.5%	8.0%	10.5%	3.8	4.7	1	0
Fermilab	15.5%	13.0%	20.5%	16.0%	3.1	2.7	0	0
CERN SPS	16.5%	15.0%	12.0%	16.0%	3.6	3.6	2	0
CERN $p\bar{p}$	1.0%	2.0%	-	2.0%	-	9.3	3	3
SLAC	3.5%	8.5%	10.0%	9.0%	4.0	4.4	1	1
DESY	7.0%	9.0%	16.5%	19.5%	7.5	8.0	6	0
Cornell CESR	1.0%	3.0%	3.0%	4.0%	18.2	9.3	2	0
Rest of world	20.0%	19.0%	11.5%	10.0%	1.4	1.3	0	0
World total	430 100%	390 100%	5190 100%	5070 100%	2.7	2.9	15	4

Conversely, if the figures for the various accelerators at CERN and the German laboratory, DESY, are combined, they suggest that the early 1980's heralded a European renaissance in high-energy physics, even before the dramatic discoveries of the W and Z particles at CERN in 1983. In particular, these two laboratories produced 80 percent of the papers cited 15 or more times in a year, while the U.S. share was just 20 percent, a complete reversal of the situation in the period 1973 to 1976 when the two European laboratories produced 25 percent of such papers and U.S. laboratories 68 percent.

How do all these scientific literature-based indicators compare with the assessments of high-energy physicists themselves? One hundred and eighty-two researchers from eleven countries were asked by us to assess six proton accelerators on a 10-point scale in terms of, one, discoveries and, two, experiments involving more precise measurements of known particles and properties. The results are given in table 8.

TABLE 8

Assessments (on a 10-point scale[1]) of main proton accelerators in terms of (a) 'discoveries' (b) providing more precise measurements

		Self-rankings	Peer-rankings	Overall rankings (sample size = 169)
Discoveries	Brookhaven AGS	9.5(±0.1)	9.0(±0.1)	9.2(±0.1)
	CERN PS	7.1(±0.2)	6.7(±0.2)	6.9(±0.1)
	CERN ISR	6.8(±0.3)	5.9(±0.2)	6.1(±0.2)
	CERN SPS	5.9(±0.3)	5.6(±0.2)	5.7(±0.1)
	Fermilab	7.4(±0.3)	7.1(±0.1)	7.2(±0.1)
	Serpukhov	3.8(±0.5)	2.6(±0.1)	2.7(±0.1)
More precise measurements	Brookhaven AGS	7.1(±0.2)	7.2(±0.2)	7.2(±0.1)
	CERN PS	8.5(±0.1)	8.5(±0.1)	8.5(±0.1)
	CERN ISR	7.3(±0.3)	6.9(±0.2)	7.0(±0.1)
	CERN SPS	8.2(±0.2)	8.2(±0.2)	8.2(±0.1)
	Fermilab	6.3(±0.2)	6.0(±0.2)	6.1(±0.1)
	Serpukhov	4.3(±0.5)	3.5(±0.2)	3.6(±0.2)

1. 10=top. The assessments are based on the relative outputs from the accelerators over their entire operational careers up to the time of the interviews with high-energy physicists in late 1981/early 1982.

Overall, we found reasonable consistency between the assessments of different groups of researchers. While there was a slight self-ranking effect, a tendency to rate one's own work more highly than others do, this was not very significant except in the case of

Serpukhov. The results show that the Brookhaven AGS was ranked well ahead of its European rival, the CERN PS, in terms of discoveries, in line with the data on highly cited papers.

Similarly, the Fermilab accelerator, which generated many more highly cited papers than the equivalent-energy CERN SPS, was ranked significantly ahead of it. In contrast, for precise measurement experiments, the PS was ranked ahead of the AGS, consistent with the data on total citations, and the SPS was likewise placed ahead of Fermilab.

Although interpretation of the picture yielded by the various indicators is by no means easy—allowance has to be made, for example, for a number of highly cited papers subsequently shown to be mistaken—there is a certain consistency between the scientific literature-based indicators and the peer rankings. It is this, together with the comments of high-energy physicists on our results, which leads us to conclude that outsiders can carry out evaluations of past research performance within individual basic science specialties. In the case of the above study, among the conclusions that we arrived at were the following:

First, with one or two prominent exceptions, nearly all the crucial discoveries in experimental high-energy physics between 1961 and 1982 were made at U.S. laboratories, even though similar energy accelerators were generally available in Western Europe. The situation has, however, changed dramatically since 1982.

Second, as regards experiments producing precise measurements and results with high statistics, the overall record of the machines at CERN taken together has been better than that of the accelerators at any laboratory in the United States.

Third, in terms of overall scientific productivity—that is, scientific performance adjusted for the differing size of the various centers, and I haven't had time to give details on this here—the record of the three U.S. national laboratories and Stanford in particular seems to have been particularly—sorry, to have been significantly better than that of CERN, their main European competitor.

What, then, is the wider significance of our approach to evaluating basic science? First, it yields information on scientific performance in a form accessible not just to researchers in the specialty concerned, but also to other scientists, science policymakers, politicians, and the public. You don't have to be a high-energy physicist to look at our tables of results and see which accelerators have been successful and which less so. It, therefore, provides a means of keeping the peer review process honest and transparent.

Second, the method raises the possibility of tracking the performance of any major research facility. This can be undertaken relatively quickly and cheaply once the initial data have been obtained. The results could help policymakers spot a significant decline in the performance of a major research center—for example, because of instrumental obsolescence—and suggest where cash infusions were most needed or, alternatively, where commitments could be reduced in order to free funds to support new areas and new people.

Last, as we describe elsewhere, the method makes possible a systematic appraisal of the prospects for a major new basic research facility compared with rival facilities worldwide. In the case of our study of world experimental high-energy physics, this involved two additional stages in the analysis.

First, we identified the factors structuring success and failure in past research performance. The most important factors reported by interviewees as accounting for the differing performance of the CERN SPS and Fermilab machines, for example, are shown in table 9.

TABLE 9

Factors explaining the relative scientific performance of Fermilab and the CERN SPS (% of interviewees believing each factor to have been important[1])

		All interviewees (n=138)	Users of Fermilab (n=39)	Users of SPS (n=66)	Users of Fermilab and SPS[2] (n=7)
1.	Fermilab had 4-year lead - SPS too late	58%	56%	64%	29%
2.	Fermilab more bold, speculative experiments — CERN solid, precise 2nd-generation experiments	39%	49%	36%	43%
3.	Problems of CERN being multinational-e.g. slow committees, over-conservative choice of experiments	11%	10%	9%	14%
4	More resources and technical support for CERN experiments - sometimes inadequate at Fermilab	50%	49%	55%	43%
5.	SPS a much better accelerator - e.g. better beams	26%	18%	32%	29%
6.	Fermilab cut too many corners building accelerator and detectors - unreliable, 'under-designed'	31%	31%	32%	29%
7.	Fermilab philosophy wrong - too much of running costs channelled into new accelerator projects	14%	18%	13%	29%
8.	Fermilab spread themselves too thinly on the ground with experiments, given the resources	13%	21%	14%	14%
9.	Energy range of both accelerators turned out to be relatively unexciting	10%	13%	8%	0%

1. These figures represent minimum values only since they are based on a content analysis of answers to a general question concerning the factors structuring the relative scientific performance of the two accelerators.

2. This is a very small sample size and the % figures may not therefore be statistically very significant.

Second, we analyzed the various factors to ascertain which are likely to continue influencing research performance in the future and to identify any new factors which may emerge over coming years.

On this basis, a set of criteria was drawn up to compare the prospects for new accelerators of the next decade, including the Stanford Linear Collider, the SLC, and the proposed Superconducting Super Collider, the SSC. Our analysis, which was carried out in 1982, led to the conclusion that the Stanford machine combined considerable scientific potential with relatively low cost. The SSC, in contrast, is an extremely expensive project, and whether its scientific potential will prove proportionately greater is by no means certain, particularly if a similar but slightly lower energy collider is completed first at CERN.

In conclusion, as industrialized nations around the world have moved into an era of approximately level science budgets, so the task of deciding which of the competing claims of researchers for new projects should be given priority has become more difficult. If policymakers were to carry out quantitative science policy studies similar to that outlined here, these would in our view provide them with information of direct relevance to this crucial task of establishing scientific priorities.

Thank you.

[The prepared statement of Professor Martin follows:]

QUANTITATIVE
SCIENCE POLICY
RESEARCH

John Irvine and Ben R. Martin*

Science Policy Research Unit
University of Sussex
Brighton BN1 9RF
United Kingdom

October 1985

To be presented to the Task Force on Science Policy of the Committee on Science and Technology, U.S. House of Representatives, 30 October 1985.

---

* No order of seniority implied (rotating first authorship). The authors are Fellows of the Science Policy Research Unit (SPRU), University of Sussex, where they work on a range of issues connected with government policies for basic and applied research. They gratefully acknowledge the financial support of the Leverhulme Trust in carrying out their current programme of research.

BEST COPY AVAILABLE

QUANTITATIVE SCIENCE POLICY RESEARCH

Introduction: general overview

Since 1978, a small team at the Science Policy Research Unit (SPRU) has carried out a range of quantitative science-policy research studies. Much of the effort has been concentrated on developing systematic methods for evaluating the research performance of laboratories, facilities and groups. Initially this work focussed on basic science, in particular on the large central facilities used to carry out research in 'big sciences' such as high-energy physics (see refs 7, 17, 18, and 19 listed in Appendix II), optical astronomy (ref. 12) and radio astronomy (ref. 13).

As we shall demonstrate below, the overall approach adopted in such institutional evaluations involves the combined use of (1) a range of science output indicators (based on publication and citation analysis) to assess the 'productivity' and 'impact' of a research facility relative to similar facilities elsewhere, and (2) the results of structured interviews with a large sample of researchers in different countries who are asked to rank these facilities in terms of their relative contributions to scientific knowledge, and to explain the factors determining differences in performance.

Over the last four years, however, the work of SPRU has diversified considerably. First, besides evaluating the outputs from research institutions, we have started to develop a programme of work on the assessment of *national* scientific performance in different fields and subfields of science, especially those with strategic technological importance for the future. This work has mainly been commissioned (by the British Advisory Board for the Research Councils and various Learned Societies) and covers such fields as protein crystallography and ocean currents (see refs 21 and 32) as well as Britain's overall scientific performance relative to other countries. Recently published results on Britain's declining scientific performance (see refs 27 and 39) attracted significant public interest in the UK earlier this year, notably during debates in the House of Lords and House of Commons.

Second, a number of studies commissioned by mainly foreign governmental agencies took us into the evaluation of *applied* R&D activities. These include projects carried out for a Norwegian Royal Commission on Industrial Research (we assessed the performance of research institutes and support mechanisms for electronics and

BEST COPY AVAILABLE

mechanical engineering - see refs 42 and 43) and for the European Commission (an evaluation of the European Community's steel research programme - see refs 45 and 47), as well as preparatory work for a current review of British support for engineering research by the Science and Engineering Research Council (ref. 50). In evaluating applied research, the approach normally involves applying a somewhat different set of indicators based upon industrial and technological impact, although publication and citation statistics were used in a recent study comparing international performance in the field of integrated optics (ref. 33).

Third, in line with the needs of policy-makers to move beyond evaluations of past performance, we have begun to explore methods for systematically appraising the future prospects of major new facilities such as accelerators (see ref. 19), work which is discussed briefly below. More importantly, the question of how one selects long-term priorities in the most effective manner was taken further in a 1984 study for the British Advisory Council on Applied Research and Development which has become interested in the question of strategic monitoring and forecasting of R&D. This study, published as a book, *Foresight in Science: Picking the Winners* (ref. 23), involved a survey of state-of-the-art methods used by governments, research-funding agencies and high-technology firms to formulate their priorities for longer-term strategic research.

Finally, over the last year we have initiated a programme of work on the *inputs* to research. The need for reliable input data (on funding, numbers of researchers, and so on) in order to understand and interpret research output data (especially those on national scientific performance in particular fields) has become ever more apparent over time. We are currently engaged in a major study to compare government funding of academic and academically related research across six countries (the United States, Japan, West Germany, France, the United Kingdom and the Netherlands). The aim is to produce not just totals for each country, but also a breakdown by 9 fields and 40 subfields. This is a very topical subject at present since senior policy-makers are increasingly interested in knowing how their spending compares with that in other countries, and the available international statistics are somewhat limited technically and in their level of disaggregation. This study is largely being financed by the British Advisory Board for the Research Councils who wish to know whether Britain's apparent decline in scientific performance is related to relative levels of funding across countries. It will be completed early in 1986.

BEST COPY AVAILABLE

In order to illustrate the type of approach used in research evaluation, we thought it would be of greatest utility to the Task Force to focus in detail upon one of our recent studies of 'big science' facilities since the question of evaluating institutional productivity, while often politically contentious, is in our view a central question facing those concerned with science policy. (Details of our other work can be found in the papers listed in Appendix II.) First, however, we should review briefly some of the main reasons behind the need for better techniques for evaluating research performance.

The need for research evaluation

In Western countries, the main decision-making mechanism in basic science continues to be peer-review - that is, relying on the opinions of scientists in the field concerned. However, peer-review faces several problems of growing importance, especially in relation to 'big science'. One stems from the trend away from the rapidly increasing national science budgets seen in the 1950s and '60s towards approximately level funding. Under the latter conditions, decisions to reduce existing financial commitments often first have to be made in order to free funds to support promising new research areas and young scientists. This is not a task for which peer-review is always very effective (ref. 16). Second, as the costs of certain research have escalated and resources have become concentrated in fewer laboratories, so it has become harder to locate 'disinterested' peers able to provide expert judgements on proposed new projects but whose own circumstances will be unaffected by the subsequent funding decision. Third, because specialties generously funded in the past are generally well represented on decision-making bodies, peer-review has exhibited a tendency to reproduce past priorities and to be less successful at identifying newly emerging (especially interdisciplinary) research areas.

Because of these problems, there is, we would argue, a need for more systematic output data to complement but not replace peer-review. To illustrate how evaluations of research output might be carried out, we describe a study which compared the scientific performance of high-energy physics accelerators around the world, including those operated by the main US laboratories (see refs 17-19, 28 and 30).

The method of converging partial indicators

How can one evaluate research performance in basic science? There are four main elements to our methodology. First, it is based on an input-output approach - it

BEST COPY AVAILABLE

involves identifying and evaluating the various inputs (funds, researchers, etc.) and outputs (for example, contributions to scientific knowledge, education, and technology), and then relating the two. For basic science, some simplification is possible since the primary output is contributions to scientific knowledge, and we therefore focus here on the evaluation of scientific contributions.

Second, the approach is institutionally focused (i.e. the unit of analysis is the research laboratory, facility, or group) because major investment decisions in basic science normally centre on institutions, and it is here that rigidities with peer-review most often arise. Third, because no absolute quantification of research performance is possible, the approach is comparative, with the proviso that one can only legitimately compare 'like' with 'like'.

Fourth, the approach involves the combined use of several indicators, each reflecting a slightly different facet of research performance. For example, publication totals (that is, the number of scientific papers published in international learned journals) give some indication of the scientific production of a research group, while numbers of papers per researcher or per dollar reveal something about its productivity. The average number of citations per paper (that is, the average number of times each paper is referred to or 'cited' by other scientists in subsequent papers) gives an indication of the impact those publications have on the scientific community, while peer-rankings (where scientists are asked to rank the relative contributions of different institutions) provide evidence on the perceived significance of the results. Last, data on the distribution of highly cited papers reveal which groups have been responsible for the few key 'discoveries' in a specialty, while citation totals reflect the large number of incremental additions to knowledge. It should be stressed that all these indicators are imperfect or 'partial' - they reflect partly the relative magnitude of contributions to scientific knowledge, and partly a variety of social, institutional, psychological, and other factors. However, when applied to matched research groups using similar research facilities, publishing in the same body of international literature subject to comparable refereeing procedures, the indicators have been found in previousstudies to yield broadly convergent results (see, for example, ref. 12). Such results we regard as being reasonably reliable - certainly more reliable than those based on a single indicator like peer-review.

The past performance of particle accelerators

Our study focussed on the world's main proton accelerators, although output data relating to electron accelerators were also produced. From Table 1, one can see that in 1959 the CERN Proton Synchrotron (PS) took over from the East European

Dubna accelerator, and before that the Berkeley Bevatron in California, as the world's highest-energy accelerator. However, the very similar and slightly higher-energy Brookhaven Alternating Gradient Synchrotron (AGS) was completed a few months later in 1960. The lead passed to the Soviet Union in 1967 when the Serpukhov accelerator was completed, but returned to the United States in 1972 when the Fermilab accelerator began operating, four years ahead of the similar SPS accelerator at CERN. Last, the CERN proton-antiproton ($p\bar{p}$) collider started operation in 1981, and this remains the world's most powerful facility to date, although the larger Fermilab Tevatron Collider is very shortly to begin experimental work.

In analyzing the outputs from the accelerators, we divided the period 1961-82 into four-year 'blocks'. Thus, Table 2 covering 1961 to 1964 shows that the Brookhaven AGS produced 4.5% of the world total of experimental high-energy physics papers over the first two years and 11% for 1963-64, some way behind its main rival, the CERN PS (29.5%), and the older Berkeley Bevatron (23%). However, the CERN papers reported data from relatively simple experiments, so their overall impact (as reflected in the number of times they were cited by other scientists) was less than that of the AGS and the Bevatron, the respective world citation-shares being 14.5%, 23% and 35% in 1964. On average, each AGS paper was cited 8.0 times in 1964, a typical figure for a new accelerator, while CERN PS publications earned only 1.9 citations per paper. Moreover, in terms of highly cited papers, the Bevatron and AGS both seem to have been responsible for more discoveries than the PS. Indeed, the three major discoveries that could have been made on either the AGS or the PS - the identification of two types of neutrinos, the omega minus, and charge-parity violation - were all made on the US accelerator.

Over the next four years, Table 3 shows that the gap between the AGS and PS narrowed. The AGS continued to yield fewer papers (19.5% compared with 25.5% in 1968), but these papers were still more highly cited (averaging 4.6 citations per paper in 1968 compared with 3.1 for CERN publications), with the result that the total impact of the two machines seems to have been similar, at least in terms of world citation-share (the figures were [illegible] and [illegible] in 1968). For papers cited 15 or more times in a year, the AGS had by then been overtaken by the PS, but the AGS continued to prove much more successful as regards the more important advances, with 12 papers cited 30 or more times compared with one from the PS.

For the period 1969-72, Table 4 reveals that, while the PS produced a steady 25% of world experimental publications, the AGS's share dropped to 14.5% in 1972 following difficulties associated with a major technical upgrade of the accelerator. However, the most highly cited papers at this time came from neither of these accelerators but from newer machines. The AGS and PS each managed only one paper cited 30 or more times, while Serpukhov, the CERN ISR and the Stanford Linear Accelerator (SLAC) each yielded four.

The next four years 1973-76 were among the most tumultuous ever in high-energy physics, ushering in the era of 'new physics'. CERN made a promising start in 1973 with the 'neutral currents' discovery and several other major advances, but after this most of the important discoveries were made in the United States. In terms of papers cited 15 times in a year, the PS and ISR were well behind SLAC and the new Fermilab accelerator near Chicago, while no less than five out of eight crucial discoveries (cited 100 times in a year) come from the SPEAR collider at Stanford. (It may be of interest to the Task Force to know that the proposal to build this highly novel and successful accelerator was turned down by the US high-energy physics community on several occasions, illustrating the conservative tendencies implicit in peer-review. As a result, Stanford had to resort to the somewhat dubious procedure of building the machine from its operating budget.) In addition, it was the Brookhaven AGS rather than the CERN PS which shared with Stanford the honour of making arguably the most important advance of the 1970s - the discovery of the J/psi particle which paved the way for the 'new physics'.

In contrast, the following four years were a period of consolidation. According to the figures in Table 6, the Fermilab accelerator seems to have contributed most during this time with about a quarter of total world citations and 40 papers cited 15 or more times. However, its performance did decline markedly towards the end of the period, partly as a result of funding problems, and partly because of the growing impact of the new CERN SPS which, taking advantage of beams and detectors significantly better than those at Fermilab, achieved a particularly high rate of citations per paper (12.7 in 1978 compared with 7.3 for Fermilab). As in previous periods, the crucial discoveries were made in the United States rather than at CERN, the two papers cited over 100 times during the four years reporting the discoveries of the upsilon at Fermilab and of parity-violation at SLAC.

AVAILABLE

Finally, what has happened since 1980? Comparison of Table 7 with earlier tables shows that Fermilab's world-share of citations had fallen from a peak of 32% in 1978 to 16% in 1982, while the figure for SLAC had likewise dropped by half between 1976 (17%) and 1982 (9%). Overall, the US share of citations has decreased sharply from 59% of the world total in 1978 to 34% in 1982. Conversely, if the figures for the various accelerators at CERN and at the German laboratory, DESY, are combined, they suggest that the early 1980s heralded a European renaissance in high-energy physics, even before the dramatic discoveries of the W and Z particles at CERN in 1983. In particular, these two laboratories produced 80% of the papers cited 15 or more times in a year, while the US share was just 20%, a complete reversal of the situation in the period 1973-76 when the two European laboratories produced 25% of such papers and US laboratories 68%.

How do all these bibliometric results compare with the assessments of high-energy physicists? 182 researchers from 11 countries were asked to assess six proton accelerators on a 10-point scale (10=top) in terms of (1) 'discoveries' and (2) experiments involving more precise measurements of known particles and properties. The results are given in Table 8. Overall, we found reasonable consistency between the assessments of different groups of researchers. While there was a slight 'self-ranking effect' - a tendency to rate one's own work more highly than do others - this was not very significant (except in the case of Serpukhov). The results show that the Brookhaven AGS was ranked well above its European rival, the CERN PS, in terms of discoveries, in line with the data on highly cited papers. Similarly, the Fermilab accelerator, which generated many more highly cited papers than the equivalent-energy CERN SPS, was ranked significantly ahead of it. In contrast, for 'precise measurement' experiments, the PS was ranked ahead of the AGS, consistent with the data on total citations, and the SPS was likewise placed ahead of Fermilab.

Although interpretation of the picture yielded by the various indicators is by no means easy (allowance has to be made, for instance, for a number of highly cited papers subsequently shown to be 'mistaken'), there is a certain consistency between the bibliometric data and the peer-rankings. It is this, together with the comments of high-energy physicists on our results, which leads us to conclude that outsiders can carry out evaluations of past research performance within individual basic-science specialties. In the case of the above study, among the conclusions that we arrived at were the following. First, with one or two prominent exceptions, nearly all the crucial discoveries in experimental high-energy physics between 1961 and 1982 were made at US laboratories, even though similar energy accelerators were generally available in Western Europe. The situation has, however, changed dramatically since 1982. Second, as regards

AVAILABLE

experiments producing precise measurements and results with high statistics, the overall record of the machines at CERN taken together has been better than that of the accelerators at any laboratory in the United States. Third, in terms of overall scientific productivity - that is, scientific performance adjusted for the differing size of the various centres (see Tables 8, 10 and 14 in reference 17) - the record of the three US National Laboratories and Stanford in particular seems to have been significantly better than that of CERN, their main European competitor.

Policy implications

What is the wider significance of our approach to evaluating basic science? First, it yields information on scientific performance in a form accessible not just to researchers in the specialty concerned, but also to other scientists, science policy-makers, politicians, and the public. It therefore provides a means of keeping the peer-review process 'honest' and transparent.

Second, the method raises the possibility of tracking the performance of any major research facility. This can be undertaken relatively quickly and cheaply once the initial data have been obtained. The results could help policy-makers spot a significant decline in the performance of a major research centre (for example, because of instrumental obsolescence), and suggest where cash infusions were most needed, or alternatively where commitments could be reduced in order to free funds to support new areas and people.

Last, as we describe elsewhere (see ref. 19), the method makes possible a systematic appraisal of the prospects for a major new basic research facility compared with rival facilities worldwide. In the case of CERN, this involved two additional stages in the analysis. First, we identified the factors structuring success and failure in past research performance. (The most important factors reported by interviewees as accounting for the differing performance of the CERN SPS and Fermilab machines, for example, are shown in Table 9.) Second, we analyzed the various factors to ascertain which are likely to continue influencing research performance in the future, and to identify any new factors which may emerge over coming years. On this basis, a set of criteria was drawn up to compare the prospects for new accelerators of the next decade, including the Stanford Linear Collider (SLC) and the proposed Superconducting Super Collider (SSC). Our analysis (which was carried out in 1982) led to the conclusion that the Stanford machine combined considerable scientific potential with relatively low cost. The SSC, in contrast, is an extremely expensive project, and whether its scientific potential will prove proportionately greater is by no means certain, particularly if a similar but slightly lower energy collider is completed first at CERN.

In conclusion, as industrialized nations have moved into an era of approximately level science budgets, so the task of deciding which of the competing claims of researchers for new projects should be given priority has become more difficult. If policy-makers were to carry out quantitative science policy studies similar to that outlined here, these would in our view provide them with information of direct relevance to this crucial task of establishing scientific priorities.

APPENDIX I

TABLES

TABLE 1

The world's main proton accelerators (>5 GeV)

Accelerator	Began operating	Beam energy (GeV)
Berkeley Bevatron (U.S.)	1954	6
JINR Dubna (E.Europe)	1957	10
CERN PS (W.Europe)	1959	28
Brookhaven AGS (U.S.)	1960	33
ITEP Moscow (U.S.S.R.)	1961	7
Argonne ZGS (U.S.)	1963	12
Rutherford Nimrod (U.K.)	1963	7
Serpukhov (U.S.S.R.)	1967	76
CERN ISR (W.Europe)	1971	31
Fermilab (U.S.)	1972	400
CERN SPS (W.Europe)	1976	400
CERN $p\bar{p}$ (W.Europe)	1981	270

TABLE 2

Experimental high-energy physics, 1961-1964

	% of papers published in past two years[2]		% of citations to work of past four years[2]	Average citations per paper	Highly cited papers: number cited n times[1]			
	1962	1964	1964	1964	n≥15	n≥30	n≥50	n≥100
Bevatron	38.0%	23.0%	35.0%	3.4	33	5	2	0
Dubna	18.5%	12.0%	3.0%	0.5	0	0	0	0
CERN PS	10.0%	29.5%	14.5%	1.9	9	1	0	0
Brookhaven AGS	4.5%	11.0%	23.0%	8.0	24	5	1	1
Moscow ITEP	2.0%	3.0%	0.5%	0.5	0	0	0	0
Rest of world	27.0%	22.5%	24.0%	2.8	22	6	1	0
World total	375 100%	535 100%	2590 100%	2.8	88	17	4	1

1. Over the period 1961-80, an experimental paper required approximately 40 or more citations in any one year to be included in the top 1% most highly cited papers for the two decades, and 19 or more to be included in the top 5%. In this respect, n≥15 corresponds to the top 7.8%, n≥30 to the top 1.9%, n≥50 to the top 0.6%, and n≥100 to the top 0.14% most highly cited papers.

2. All figures in this and subsequent tables have been rounded to the nearest 0.5%.

[illegible] AVAILABLE

TABLE 3

Experimental high-energy physics, 1965-1968

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers: number cited n times			
	1966	1968	1966	1968	1966	1968	n≥15	n≥30	n≥50	n≥100
Bevatron	12.5%	11.5%	17.5%	10.0%	3.8	2.8	9	0	0	0
Dubna	6.5%	4.0%	2.5%	1.0%	1.1	0.9	0	0	0	0
CERN PS	32.5%	25.5%	28.5%	26.0%	3.5	3.1	46	1	0	0
Brookhaven AGS	19.5%	19.5%	28.0%	26.5%	6.8	4.6	42	12	1	0
Argonne ZGS	2.0%	6.5%	1.0%	5.0%	4.2	3.6	6	0	0	0
Rutherford Nimrod	2.5%	2.5%	2.5%	3.5%	6.0	4.6	9	0	0	0
Rest of world	24.0%	30.5%	20.0%	28.0%	3.3	3.4	38	5	0	0
World total	645 100%	845 100%	4500 100%	5080 100%	3.8	3.4	150	18	1	0

BEST COPY AVAILABLE

TABLE 4

Experimental high-energy physics, 1969-1972

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers : number cited n times			
	1970	1972	1970	1972	1970	1972	n≥15	n≥30	n≥50	n≥100
Bevatron	9.0%	6.0%	10.0%	6.5%	2.9	2.5	3	1	0	0
CERN PS	25.0%	25.0%	22.5%	21.5%	2.6	2.3	19	1	0	0
Brookhaven AGS	19.0%	14.5%	23.0%	18.0%	3.5	2.9	22	1	0	0
Argonne ZGS	9.5%	9.5%	7.5%	9.0%	2.8	2.6	7	0	0	0
Rutherford Nimrod	2.5%	2.5%	2.5%	2.0%	3.2	1.9	0	0	0	0
Serpukhov	3.0%	6.0%	2.5%	5.0%	4.8	3.0	12	4	2	1
CERN ISR	-	2.0%	-	5.0%	-	12.9	13	4	2	0
SLAC	5.5%	8.0%	8.5%	12.5%	6.1	4.9	21	4	2	0
Rest of world	27.0%	26.0%	23.0%	19.5%	2.4	2.0	21	6	3	0
World total	950 100%	1020 100%	5270 100%	5310 100%	2.9	2.7	123	21	9	1

BEST COPY AVAILABLE

TABLE 5

Experimental high-energy physics, 1973-1976

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers : number cited n times			
	1974	1976	1974	1976	1974	1976	$n \geq 15$	$n \geq 30$	$n \geq 50$	$n \geq 100$
CERN PS	20.0%	23.5%	19.5%	14.5%	2.7	2.2	18	4	2	0
Brookhaven AGS	13.5%	8.0%	11.0%	10.0%	2.5	3.2	15	5	2	1
Argonne ZGS	9.0%	6.0%	7.0%	4.5%	2.3	2.1	5	0	0	0
Rutherford Nimrod	3.0%	1.5%	2.0%	0.5%	2.1	1.1	0	0	0	0
Serpukhov	9.5%	10.5%	7.0%	6.5%	2.8	2.1	7	0	0	0
CERN ISR	3.5%	5.0%	12.0%	9.0%	14.0	7.4	28	12	2	1
Fermilab	9.5%	15.0%	11.5%	25.0%	6.9	6.7	71	26	9	0
SLAC	7.0%	9.0%	10.0%	17.0%	4.2	7.1	37	14	8	5
Rest of world	26.0%	21.5%	20.0%	12.5%	2.4	1.8	16	6	3	1
World total	1150 100%	1180 100%	6780 100%	7740 100%	3.1	3.3	197	57	26	8

TABLE 6

Experimental high-energy physics, 1977-1980

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers: number cited n times			
	1978	1980	1978	1980	1978	1980	n≥15	n≥30	n≥50	n≥100
CERN PS	22.0%	11.5%	14.5%	12.5%	2.2	2.2	13	2	1	0
Brookhaven AGS	5.5%	5.5%	5.0%	3.0%	2.7	1.6	0	0	0	0
Serpukhov	12.0%	14.0	4.0%	5.0%	1.2	1.2	0	0	0	0
CERN ISR	4.5%	5.5%	7.0%	7.5%	5.4	4.4	11	2	0	0
Fermilab	16.5%	19.0%	32.0%	21.5%	7.3	3.6	40	10	5	1
CERN SPS	2.5%	8.5%	4.0%	8.5%	12.7	5.0	19	7	3	0
SLAC	9.5%	6.0%	15.0%	11.5%	5.7	4.4	25	6	1	1
DESY	4.0%	6.5%	5.5%	15.5%	5.7	8.8	36	16	4	0
Rest of world	23.5%	24.0%	13.0%	15.0%	2.0	1.9	19	5	0	0
World total	1115 100%	930 100%	8190 100%	6090 100%	3.5	3.0	164	48	14	2

TABLE 7

Experimental high-energy physics, post-1980

	% of papers published		% of citations to work of past four years		Average citations per paper		Highly cited papers : number cited n times	
	1981	1982	1981	1982	1981	1982	n≥15	n≥30
CERN PS	9.5%	7.0%	8.5%	5.0%	1.6	1.4	0	0
Brookhaven AGS	4.5%	3.0%	3.0%	2.5%	1.7	1.4	0	0
Serpukhov	15.0%	11.5%	6.5%	6.0%	1.4	1.3	0	0
CERN ISR	6.0%	9.5%	8.0%	10.5%	3.8	4.7	1	0
Fermilab	15.5%	13.0%	20.5%	16.0%	3.1	2.7	0	0
CERN SPS	16.5%	15.0%	12.0%	15.0%	3.6	3.8	2	0
CERN $p\bar{p}$	1.0%	2.0%	-	2.0%	-	9.3	3	3
SLAC	3.5%	8.5%	10.0%	9.0%	4.0	4.4	1	1
DESY	7.0%	9.0%	16.5%	19.5%	7.5	8.0	6	0
Cornell CESR	1.0%	3.0%	3.0%	4.0%	18.2	9.3	2	0
Rest of world	20.0%	19.0%	11.5%	10.0%	1.4	1.3	0	0
World total	430 100%	390 100%	5190 100%	5070 100%	2.7	2.9	15	4

TABLE 8

Assessments (on a 10-point scale[1]) of main proton accelerators in terms of (a) 'discoveries' (b) providing more precise measurements

		Self-rankings	Peer-rankings	Overall rankings (sample size = 169)
Discoveries	Brookhaven AGS	9.5(±0.1)	9.0(±0.1)	9.2(±0.1)
	CERN PS	7.1(±0.2)	6.7(±0.2)	6.9(±0.1)
	CERN ISR	6.8(±0.3)	5.9(±0.2)	6.1(±0.2)
	CERN SPS	5.9(±0.3)	5.6(±0.2)	5.7(±0.1)
	Fermilab	7.4(±0.3)	7.1(±0.1)	7.2(±0.1)
	Serpukhov	3.8(±0.5)	2.6(±0.1)	2.7(±0.1)
More precise measurements	Brookhaven AGS	7.1(±0.2)	7.2(±0.2)	7.2(±0.1)
	CERN PS	8.5(±0.1)	8.5(±0.1)	8.5(±0.1)
	CERN ISR	7.3(±0.3)	6.9(±0.2)	7.0(±0.1)
	CERN SPS	8.2(±0.2)	8.2(±0.2)	8.2(±0.1)
	Fermilab	6.3(±0.2)	6.0(±0.2)	6.1(±0.1)
	Serpukhov	4.3(±0.5)	3.5(±0.2)	3.6(±0.2)

1. 10=top. The assessments are based on the relative outputs from the accelerators over their entire operational careers up to the time of the interviews with high-energy physicists in late 1981/early 1982.

BLANK PAGE

BLANK PAGE

BLANK PAGE

TABLE 9

Factors explaining the relative scientific performance of Fermilab and the CERN SPS (% of interviewees believing each factor to have been important[1])

		All interviewees (n=130)	Users of Fermilab (n=39)	Users of SPS (n=56)	Users of Fermilab and SPS (n=7)[2]
1.	Fermilab had 4-year lead - SPS too late	58%	56%	54%	29%
2.	Fermilab more bold, speculative experiments — CERN solid, precise 2nd-generation experiments	39%	49%	36%	43%
3.	Problems of CERN being multinational - e.g. slow committees, over conservative choice of experiments	11%	10%	9%	14%
4	More resources and technical support for CERN experiments - sometimes inadequate at Fermilab	50%	49%	55%	43%
5.	SPS a much better accelerator - e.g. better beams	26%	15%	32%	29%
6	Fermilab cut too many corners building accelerator and detectors - unreliable, 'under-designed'	31%	31%	32%	29%
7	Fermilab philosophy wrong - too much of running costs channelled into new accelerator projects	14%	18%	13%	29%
8	Fermilab spread themselves too thinly on the ground with experiments, given the resources	13%	21%	14%	14%
9	Energy range of both accelerators turned out to be relatively unexciting	10%	13%	5%	0%

1. These figures represent minimum values only since they are based on a content analysis of answers to a general question concerning the factors structuring the relative scientific performance of the two accelerators.

2. This is a very small sample size and the % figures may not therefore be statistically very significant.

APPENDIX II

PUBLICATION LIST

Publications and Reports from the Science Policy and Research Evaluation Group, Science Policy Research Unit

(a) *Basic Research*

1. N. Dombey, J. Irvine, B. R. Martin, R. Turner and K. Pavitt, 1980, *Final Report of the Project on 'Criteria for Capital Projects in Basic Science'*, London: Social Science Research Council.

2. J. Irvine and B. R. Martin, 1980, 'The Economic Effects of Big Science: The Case of Radio Astronomy', *Proceedings of the International Colloquium on Economic Effects of Space and Other Advanced Technologies, Strasbourg 28-30 April 1980*, Paris: European Space Agency (Ref. ESA SP-151, September 1980).

3. J. Irvine and B. R. Martin, 1980, 'A Methodology for Assessing the Scientific Performance of Research Groups', *Scientia Yugoslavica*, 6 (1-4), pp.83-95.

4. B. R. Martin and J. Irvine, 1980, 'Output Indicators for Basic Research: Some Policy Tools for Assessing Scientific Progress', OECD Conference on Science and Technology Indicators, September 1980, Paris: Organization for Economic Co-operation and Development (Ref. STIC/80.37).

5. B. R. Martin and J. Irvine, 1981, 'Spin-off from Basic Science: The Case of Radio Astronomy', *Physics in Technology*, 12, pp.204-12.

6. J. Irvine and B. R. Martin, 1981, 'L'Evaluation de la Recherche Fondamentale: Est-Elle Possible?', *La Recherche*, 12 (128), pp.1406-16.

7. B. R. Martin and J. Irvine, 1981, 'Internal Criteria for Scientific Choice: An Evaluation of the Research Performance of Electron High-Energy Physics Accelerators', *Minerva*, XIX(3), pp.408-32.

8. J. Irvine and B. R. Martin, 1982, 'Es Posible Valorer la Investigacion Pura?' *Mundo Cientifico*, 2 (11), pp.184-95.

9. B. R. Martin and J. Irvine, 1982, 'Women in Science: The Astronomical Brain Drain', *Women's Studies International Forum*, 5 (1), pp.41-68.

10. J. Irvine, B. R. Martin and G. Oldham, 1982, *Research Evaluation in British Science: A SPRU Review*, Report to *French Ministry of Industry and Research, Paris, France*.

11. K. L. R. Pavitt, B. R. Martin and J. Irvine, 1982, *Final Report of the Project on 'CERN: Past Performance and Future Prospects'*, London: Social Science Research Council.

12. J. Irvine and B. R. Martin, 1983, 'Assessing Basic Research: The Case of the Isaac Newton Telescope', *Social Studies of Science*, 13 (1), pp.49-86.

13. B. R. Martin and J. Irvine, 1983, 'Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, 12 (2), pp.61-90.

14. J. Irvine and I. Miles, 1983, 'Science and Technology as Critical Issues for Alternative Ways of Life', in A Sicinski and M. Wemegah (eds.), *Alternative Ways of Life in Contemporary Europe*, Tokyo: UNU, pp.157-71.

15. J. Irvine and B. R. Martin, 1983, 'The Isaac Newton Telescope', *Social Studies of Science*, 13, 321-22.

16. J. Irvine and B. R. Martin, 1984, 'What Direction for Basic Scientific Research?', in M. Gibbons, P. Gummett and B. M. Udgacnkar (eds.), *Science and Technology Policy in the 1980s and Beyond*, Harlow: Longman, pp.67-98.

17. B. R. Martin and J. Irvine, 1984, 'CERN: Past Performance and Future Prospects - I - CERN's Position in World High-Energy Physics', *Research Policy*, 13, pp.183-210.

18. J. Irvine and B. R. Martin, 1984, 'CERN: Past Performance and Future Prospects - II - The Scientific Performance of the CERN Accelerators', *Research Policy*, 13, pp.247-34.

19. B. R. Martin and J. Irvine, 1984, 'CERN: Past Performance and Future Prospects - III - CERN and the Future of World High-Energy Physics', *Research Policy*, 13, pp.311-42.

20. S. Plummer, B. R. Martin and J. Irvine, 1984, 'L'Evaluation de la Recherche: Resultats du Radiotelescope de Nancay', *La Recherche*, 15(161), pp.1610-11.

21. B. R. Martin, J. Irvine and D. Crouch, 1984, *Bibliometric Analyses of Ocean Currents and Protein Crystallography: A Report for the ABRC Science Policy Study*, mimeo, London: ABRC.

22. J. Irvine and B. R. Martin, 1984, *Project Foresight: An Assessment of Approaches to Identifying Promising Areas of Science*, Report to the Cabinet Office/ACARD by the Science Policy Research Unit, mimeo, Brighton: SPRU.

23. J. Irvine and B. R. Martin, 1984, *Foresight in Science: Picking The Winners*. London: Frances Pinter (Publishers).

24. J. Irvine, 1984, 'Evaluating Basic Research: The Case of CERN', *Proceedings of the Nordic Scientific Policy Council Conference on Research Evaluation, Helsinki, February, 1984*, Copenhagen: NSPC.

25. B. R. Martin and J. Irvine, 1984, 'Research Evauation. Why? How?' in W. Callebaut, S. Cozzens, B.-P. Lecuyer, A. Rip and J.-P. Van Bendegem (eds.), *George Sarton Centennial*, Ghent, Belgium: Communication and Cognition, pp.241-43.

AVAILABLE

26. J. Irvine, B.R. Martin and G. Oldham, 1984, 'Evaluation des Recherches Menees Par la Science Britannique: Un Rapport du SPRU', *CPE Etude,* 34, pp.143-84.

27. B. R. Martin, J. Irvine and R. Turner, 1984, 'The Writing on the Wall for British Science', *New Scientist*, 104, October 8, pp.25-29.

28. J. Irvine and B. R. Martin, 1985, 'Evaluating Big Science: CERN's Past Performance and Future Prospects', *Scientometrics*, 7 (3-6), pp.281-308.

29. J. Irvine and B. R. Martin, 1985, 'Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators', *Social Studies of Science*, 15, pp.293-341.

30. B. R. Martin and J. Irvine, 1985, 'CERN's Past Scientific Performance: A Summary Review', *Nature* (forthcoming).

31. B. R. Martin, J. Irvine and D. Crouch, 1985, *Science Indicators for Research Policy: A Bibliometric Analysis of Ocean Currents and Protein Crystallography*, SPRU Occasional Paper No.23, Brighton: University of Sussex.

32. D. Crouch, J. Irvine and B.R. Martin, 1985, 'Bibliometric Analysis for Science Policy: An Evaluation of the United Kingdom's Research Performance in Ocean Currents and Protein Crystallography', *Scientometrics* (forthcoming).

33. D. Hicks, B.R. Martin and J. Irvine, 1985, 'Bibliometric Techniques for Monitoring Performance in Strategic Research: The Case of Integrated Optics', in submission to *R&D Management*.

34. B. R. Martin and J. Irvine, 1985, 'Evaluating the Evaluators: A Reply to Our Critics', *Social Studies of Science*, 15, pp.558-75.

35. B. R. Martin and J. Irvine, 1985, 'Research Evaluation in Big Science: A Case-Study of CERN', in G. Statera and L. Cannavo (eds.), *Proceedings of the International Conference on Scientific Professionalism, Rome, 20-23 March 1985, Volume I: Post-Industrial Society and Research Policy*, Milan: Franco Angeli (forthcoming).

36. B. R. Martin and J. Irvine, 1985, 'Output Indicators for Science Policy: An Evaluation of CERN's Past Performance and Future Prospects, paper presented at Workshop on Science and Technology Indicators, OECD, 10-13 June, 1985, DSTI/SPR/85.24/17A, Paris: OECD.

37. J. Irvine and B. R. Martin, 1985, 'Women in Radio Astronomy', paper presented at the Annual Meeting of the British Association for the Advancement of Science, University of Strathclyde, Glasgow, 26-30 August 1985; to be published in the proceedings.

38. B. R. Martin and J. Irvine, 1985, 'Uses of Citation Analysis: An Assessment of CERN', paper presented at the Annual Meeting of the British Association for the Advancement of Science, University of Strathclyde, Glasgow, 26-30 August 1985.

39. J. Irvine, B.R. Martin, T. Peacock and R. Turner, 1985, 'Charting the Decline in British Science', *Nature*, 316 (6029), pp.587-90.

40. B.R. Martin, J. Irvine, T. Peacock and J. Abraham, 1985, 'A Re-Evaluation of the Contributions to Radio Astronomy of the Nancay Observatory', paper presented at Annual Conference of the Society for Social Studies of Science, Troy, New York, October; to be submitted to *4S Review*.

41. J. Irvine, B. R. Martin, J. Abraham and T. Peacock, 1985, 'Assessing Basic Research: Reappraisal and Update of an Evaluation of Four Radio Astronomy Observatories', to be submitted to *Research Policy*.

(b) *Applied Sciences and Engineering*

42. M. Schwarz, J. Irvine, B. R. Martin, K. Pavitt and R. Rothwell, 1982, Government Support for Industrial Research: Lessons from a Study of Norway', *R&D Management*, 12 (4), pp.155-67.

43. J. Irvine, B. R. Martin and M. Schwarz, with K. Pavitt and R. Rothwell, 1961, *Government Support For Industrial Research in Norway: A SPRU Report*, (Norwegian Official Publication NOU 1981: 30B), Oslo: Universitetsforlaget.

44. B. R. Martin and J. Irvine, 1983, *Perspectives on the Norwegian Institute for Energy Technology*, Oslo: Ministry of Oil and Energy (restricted).

45. J. Irvine and B. R. Martin, 1983, *The Economic and Social Impact of the ECSC Steel Research Programme: A SPRU Evaluation*, Brussels: Commission of the European Communities (restricted).

46. B. R. Martin, 1984, 'Evaluating Applied Research: The Case of Norwegian State Support for Industrial R&D', *Proceedings of the Nordic Scientific Policy Council Conference on Research Evaluation, Helsinki, February 1984*, Copenhagen: NSPC.

47. B. R. Martin, J. Irvine and J. Aylen, 1984, 'Evaluating International Collaboration in R&D: The Case of Steel Research in the European Community', to be submitted to *R&D Management*.

48. B. R. Martin and J. Irvine, 1984, *Assessment of Priorities for Building and Civil Engineering Research*, Report to the Department of Environment and NEDO, mimeo: SPRU (restricted).

49. B. Martin, 1985, 'British Research - Decline and Fall? And What Can We Do About It?', in D. Gregory (ed.), *Proceedings of the Technical Futures Symposium on Building Services in the Year 2000, Sunningdale, 19-21 April 1985*, Bracknell, Berkshire: BSRIA.

50. J. Irvine and B. R. Martin, 1985, *Assessing the Impact of SERC Support for Engineering Research*, report to the Science and Engineering Research Council, mimeo, Swindon: SERC (restricted).

AVAILABLE

DISCUSSION

Mr. Fuqua. Thank you very much, Professor Martin.

From your statement, you indicated that a significant part of your work has been supported by the Research Councils in Great Britain.

To what extent to date have they found actual use and application in the formulation of their own science policies?

Professor Martin. Initially there was a lot of skepticism to the very first studies that we did starting in 1977-78 and when those first results came out I am not sure exactly how much influence they had. There was certainly a lot of controversy with some senior scientists in particular, if not disputing the results, at least disputing any interpretation on them that we might have placed.

However, I think the situation did begin to change with the study that I reported here. When we finally produced the results of that study and showed them to senior policymakers—there was one in particular who had played a major role in discussing Britain's contribution or participation in CERN and, in particular, in the new LEP project which is current being built at CERN. He, when writing to us after he received these papers, said that he found the results that we produced very useful. The only pity was that they hadn't come a few months earlier, because they might then have been even more useful. There had been a delay in getting the funding for this project.

Mr. Fuqua. There was a recommendation by the British Government that they reduce their effort in CERN by 25 percent. Did your study contribute to that decision?

Professor Martin. Yes. No. That committee came a year or so after we finished this study. We submitted the results of our study to that committee, the Kendrew Committee, and we also made a presentation in front of the committee.

Again, I don't know exactly how much influence our findings may have had on the committee, but again we've had comments to the effect that at least people were able to see more clearly what was happening within high-energy physics. They didn't have to rely on high-energy physicists to tell them how successful CERN had been in the past or to look at the comparative strengths and weaknesses of the big new machine that is being built at CERN at present.

Mr. Fuqua. I notice in your statement you have been doing this work for Norway. Have they utilized your results?

Professor Martin. Yes. As I said, when we began our work there was a lot of skepticism toward whether or not it was useful for policymaking purposes. That skepticism was particularly pronounced in Britain. We spent the first three or four years of our research trying to convince policymakers in Britain that we had something to offer which might be of relevance to their problems.

Our first break came not from a British organization, but from Norway, from a Royal Commission set up to look at the mechanisms in Norway for supporting applied R&D. What we had to do was visit a number of the collective research institutes which are operated by the Norwegian Council for Applied Research and look at the work which was carried out at that institute which is supposed to provide longer term research of relevance to the needs of industry.

We concentrated, as I said, on two areas, mechanical engineering, an old traditional area, and electronics, a new, fast-moving area, and by interviewing the researchers at those institutes and also interviewing R&D managers in the two industrial sectors concerned, we were able to build up a picture as to the utility of the research which was going on in those institutes and how successful it was from the point of view of industry.

We then produced our report which went to the Royal Commission and we were told that the Royal Commission's report was based very heavily on ours. It's in Norwegian, so we weren't able to establish that exactly. And then, in turn, a white paper was drawn up which apparently was based very heavily on what the Royal Commission had reported.

Now I suppose the best test of how useful that work was is—it's a bit like politicians: do you get invited back a second time? We were invited back a second time about a year later to do another study in Norway, this time on a research institute which was formerly an atomic energy institute but which, because Norway has made a political decision not to build any commercial power reactors, at least until the next century, they don't need a big institute just doing atomic energy research and, therefore, it's being asked to diversify into other areas of R&D, and it hasn't been—at that stage hadn't been that successful.

Questions were asked in their parliament, and a committee was set up to look at this process of diversification at the institute. The committee wasn't getting very far, and suddenly we got a phone call saying, could we come back and within 6 weeks produce a report based on what was happening in this institute and also in four other similar institutes in other European countries, which had faced this same problem of diversifying away from atomic energy into more industrially relevant R&D.

So, as I say, the best test that we can think of as to whether our results are useful is whether we are invited back a second time, and we were in that case.

Mr. Fuqua. Let's get on another subject to see if you get invited back. [Laughter.]

On page 3 of your prepared statement, middle of the page, on the need for research evaluation, you cite three issues relating to peer review. Peer review probably has been the principal way of obtaining evaluations of research for some time. It is no secret that scientists prefer this as a method of evaluating proposals rather than a quantitative approach such as that you have taken.

We have received testimony from other people which has suggested that we should look at a researcher's track record. What have they published? What is his track record as it relates to this particular area?

Have you been able to—have you been invited back to—overcome this skepticism that exists in the scientific community about the pure aspects of peer review?

Professor Martin. Certainly there is, or was, a lot of skepticism, as I say, in the early days when we were doing this, and not just

AVAILABLE

skepticism—outright hostility or refusal to admit that our approach had anything useful to offer.

When we finished that first study and went on to the study of CERN, there was a question raised, which is: Would high-energy physicists in our study of CERN be willing to talk to us? Knowing that we had already done some of these studies on astronomy, regular astronomy and optical astronomy, would they be willing to talk to us freely and frankly, as they had been in those earlier studies, now that they were aware of the sorts of use to which we could put the results of our studies?

So, to a certain extent, there was an experiment involved here, and we found that in one or two cases, when we went to visit high-energy physicists and asked them to talk about their own work, how successful it had been, and then how successful their centers compared to other centers, they knew what we were getting at, which was to try and rank centers and perhaps their center wasn't doing very well.

So perhaps, initially, in some cases there was, if not hostility, a little bit of reticence on their part, not willing to discuss this. But, by and large, we were able to convince the great majority of the ones that we interviewed that the approach did have some utility and that some validity, and they were in most cases still able to give us their views as to how successful different accelerators would be. So we were able to do it a second time.

There is another question, though, and that is the extent to which, once people see the indicators we are using, will they then attempt to start manipulating them? Will they start trying to produce more papers in learned journals by splitting papers into two? Will they try to get more citations by doing perhaps more controversial work or by getting their friends to cite their papers?

So one thing that we're currently doing is going back to the very first study we did—which was radio astronomy, and we are updating that study, adding on another five years—to see, first of all, what has happened in the intervening five years, but, secondly, to try to explore the extent to which scientists have or have not attempted to start manipulating these indicators to their own advantage.

We suspect that you may be able to do it to a small extent, but word will quickly get around that the laboratory is splitting all of its papers into small papers or if it's engaging in shady citation practices, and this will be heavily discounted when you come to talk to the scientists themselves about what have been the more successful laboratories.

So we think that the method can be used over the longer term. It's not a one-time approach.

Mr. Fuqua. I have just one other question and then I will yield to Mr. Lujan.

You cite throughout your testimony the evaluation of the Federal laboratories, which is a rather broad sample because of the many people involved in each of those that you named. That makes it probably a more accurate conclusion than, say, where you're doing the individual investigators that may be located at a university. You mentioned one in radio astronomy, or physics, or chemistry, or mathematics, or some of those other fields of science.

How much experience do you have in those fields before you evaluate an individual researcher?

Professor Martin. In none of our studies have we focused on individual researchers. The main reason for this is that we see the main problems from a science policymaking point of view focusing on large groups, large institutions, and so on. This is particularly true in the United Kingdom where our science budget, the money that goes to the research council, is very heavily concentrated on a few laboratories.

If you take the biggest council, the Science and Engineering Research Council, about 60 or 70 percent of the money it spends on science goes on certain laboratories. So from a policymaking point of view, we think it's most intriguing to focus on some of these big facilities.

We have, therefore, attempted to devise indicators that relate to such large groups of facilities. We are aware that, once one starts to try and apply the same indicators to individuals, you run into a whole host of problems, and I personally would be unwilling to apply these sorts of indicators to individuals. There's too much individual variation in publication practices and referencing habits of scientists in my view to make this valid for individual researchers unless handled with very great care; in other words, with a lot of peer review.

Mr. Fuqua. It would be better with a larger group such as laboratories?

Professor Martin. It is, because to some extent you can then choose comparable groups with, for example, optical telescopes, you can look at similar-sized telescopes, with accelerators; you can look at similar energy telescopes. So it's possible to find matched groups that you can apply these to, perhaps more accurately than one could with individuals where everybody's doing slightly different things.

Mr. Fuqua. Mr. Lujan?

Mr. Lujan. Thank you very much, Mr. Chairman.

It's really quite interesting testimony. It will take some time to digest, and I intend to do that. But it is particularly relevant I think to our task force in studying science policy.

One thing that stands out as you go through your testimony is that certain laboratories, say in the accelerator business, rate high on your rating system at one time, and then it's replaced by someone else. It's not unlike—my background is a business background—in the service industry you find that very much. I've never known the reason for it, except that somebody gets self-satisfied that "we're the leader."

Now in transportation, some transportation company will be the big one. In a bank, in a travel agency, real estate agency—those kinds of service businesses, you notice one company for a period of time dominating the scene and then moving on. So it's not so unusual in science.

Do you think that will continue and does the fact that someone makes a discovery just spur the other one to a little more activity? Do you think studies are partially responsible for the lead moving from laboratory to laboratory?

AVAILABLE

Professor **Martin**. OK. I think the best way of answering that question is by way of concrete example. I mentioned the study that we did in Norway, the first time we had been commissioned by a policymaking body to carry out a study. In that study we looked at a number of institutes, two in particular, the two main applied research institutes in Norway, one in Oslo and one in Trondheim.

We found that both those laboratories, which were set up in the late forties and early fifties, had initially been very successful. They had good researchers; they did some good projects which were very—had a big impact on industry.

However, over the last 10, 15 years one of those two laboratories has—the process or what you have just described has taken place; a certain amount of stagnation has taken place in that laboratory, whereas the other one has continued to be very successful both as regards to quality of its research and its impact on industry. So the interesting question was why had these two laboratories which initially had been fairly similar—one had continued to be successful and the other less so.

It looked as though one of the important factors here was that one of those laboratories in effect operated in isolation. The other one integrated into a technical university, and what is important in that second case is the interaction you get between the research laboratory and the teaching. It works in a number of ways. A laboratory that's integrated into a teaching university or environment is able to recruit good students, and therefore you get a good inflow of bright young people.

Another interaction is that the researchers, many of them have to do a certain amount of teaching. That is a challenging process. They have to read more widely than they would otherwise do. Research tends to have a specializing trend built in it, and that is offset to some extent if you are forced to teach. So that second institute, because of having this continuous inflow of new people and new ideas, appears—those appear to be some of the factors explaining why it's been able to resist the complacency which might otherwise have set in.

Now—so that was the situation when we did the study. The institute that we thought that wasn't doing very well when we produced our report was obviously a little bit unhappy with the results we came to and implied that perhaps we didn't have the competence as outsiders to make judgments on them.

It's interesting, though, that when I went back to Scandanavia 2 or 3 months—sorry, 2 or 3 years later to attend a conference, and we were a little bit worried when we saw the list of attendees, people attending the conference. One of them was the director of this institute, who had been rather rude to us at the end of the study.

But, anyway, we arrived at the conference, turned up, and along came the director to speak to us, and he was smiling. And what had happened was that he had, first of all, once he'd accepted that our report had some validity—and that took him a little while to get used to—he then used the results of our report—and we were outsiders, we had no axe to grind in the Norwegian situation—to force through changes in his laboratory that would otherwise have been rather hard to make. And, therefore, he has apparently been able to change the performance of this laboratory.

And just one little indication of this—apparently that year that we spoke to him, or the year before perhaps, the laboratory had been so successful it had made more profit from the work that it did for industry, and all the staff had been given a Christmas bonus, which is unheard of apparently in such Norwegian institutes.

So one can—well, first of all, yes, you can fall victim to this sort of complacency you were mentioning, but there are ways of offsetting that.

Mr. Lujan. There's one concern that we have had in this committee, at least some of us have had in this committee, of the rich get richer. Your method would seem to strengthen that—the rich getting richer process, because if some new university or new laboratory has not had very good success, if we follow your method of funding—you can overcome that by considering the fact we have a big emphasis on training scientists. That is one of the criteria that we use. How can we overcome that, big laboratories getting all the money and the little ones not?

Professor Martin. OK. I would argue that if one relies just on peer review there is equally a case that the rich will get richer.

Mr. Lujan. Yes.

Professor Martin. And we've seen various examples of this where the poor, the poorly represented, the smaller laboratories, because they are not well represented on decisionmaking bodies, don't get their new ideas funded. And one good example of this was the SPEAR accelerator at SLAC. At the time, SLAC did not have a large outside user group in the way that the two other big U.S. laboratories did, and this may have been one reason why. Although that looked a very good project and subsequently turned out to be perhaps the most successful accelerator ever built, it was turned down on three occasions by the high-energy physics community at large.

So I would argue that peer review is just as much subject to this problem of the rich getting richer.

Yes, if one just looked at past performance, then this might be a problem with our indicators, and it was one that was raised when we finished our first big science study in 1980. People said, "OK, you told us about past performance but we're concerned more with future prospects."

So this is why in the next study, the study of world high-energy physics, we did two extra things. One is to look at the factors structuring performance in the past, and these were identified mainly by interviewing researchers in the field, again with a structured interview and set of questions. And on the basis of this, we got some feeling for what sorts of factors account for success and failure among accelerators.

Then the second stage, as I mentioned earlier, was to look at those factors more closely and identify which of them were likely to continue influencing things in the future, which ones might drop out, what new ones might emerge. And on the basis of that, I would argue that it is possible to systematically appraise the pros-

AVAILABLE

pects for new accelerators in a way that does not necessarily lead to the rich getting richer.

If I just give one example, one of the laboratories that we looked at is the German National Laboratory, DESY, which is by no means one of the largest accelerator laboratories in the world. Yet, the machine that they're proposing to build, an electron-proton collider, looks, on the basis of the criteria which we drew up as to factors that are likely to structure performance in the future, as though it might be potentially the most interesting machine to build at the present.

So I would suggest that, if you look at factors structuring performance, one would get around this tendency to just allow the rich to get richer.

Mr. LUJAN. One final question—really two questions which are related:

To your knowledge, are there any individuals or groups or Government agencies here in the United States where your methodology is now being tried to evaluate science projects? That's the first part of the question. The second is, if not, what kind of staff do we need to bring that about?

Professor MARTIN. The people doing the nearest equivalent sort of work to our own I think are to be found in the National Science Foundation and in NIH, where over several years now they have either been carrying out studies or commissioning studies which use, to a greater or lesser extent, some of these quantitative indicators relating to science policy.

As far as we know, there is no one in the States who has done exactly the sort of studies we have focusing on big facilities in particular. However, there has been a lot of work looking at national performance both overall and in individual fields. I am thinking in particular of the very interesting work that is being done by Dr. Narin at Computer Horizons, Inc. He produces data every 2 years for the National Science Foundation's Science Indicators Report, which we have found very useful. We have, in fact, purchased copies of the data that he produces and used it to look at British performance. So there is that work going on, looking at national performance. I'm less aware of work looking at institutional performance.

Mr. LUJAN. What kind of staffing—assuming you want to evaluate our national laboratories—we fund our laboratories—as a matter of fact, our entire research in our national laboratories—in a competitive manner. They compete for the best ways by using different means.

If you wanted to evaluate those to see what their output has been over the last 20 years, 30, or 40 years, comparing one against the other, what kind of staffing would be required, the size of the staff?

Professor MARTIN. The most labor-intensive part of our work is compiling the initial data on publications in the scientific literature, and the number of times they're cited, and so on. In fact, one of the main problems is insuring that you have complete publication lists, and in many cases laboratories, for one reason or another, don't have a complete list and we have to try to devise ways to make those lists more complete.

AVAILABLE

In the case of the high-energy physics study, for example, although the laboratories do keep their own records, we have found that certainly in the past those records were not at all complete and we, therefore, had to resort to a procedure which involves scanning the 11 leading international journals in the field and picking out all the experimental high-energy physics papers from them.

So there is a lot of work there, but that could be reduced if procedures were set up by funding agencies to insist that much more comprehensive records were kept of the published output from the different facilities. And if you have a multifacility laboratory, then those publications should be sorted according to the facility on which the work was done.

That would then greatly reduce the amount of work that would need to be done in studies like these, and then one might be talking about a fairly small staff. I wouldn't like to put exact numbers on this. Just to give you some idea, the work that I've been describing is being done essentially by John Irvine and myself aided by research assistants. And, well, we find that, say a study of high-energy physics worldwide, that took about two to three man-years. That is the sort of effort that is involved, given that there wasn't very good data to begin with on publications.

So I would have thought that a staff somewhere on the order of 6 to 10 people would be sufficiently large to do a range of studies each year, looking at some of the areas where there was a particular policy issue or problem being considered at about that time.

Mr. LUJAN. It would be much easier if you took one field? Say you wanted to know who has done what or the primary work or have you really gotten results in, say, fusion, for example.

Professor MARTIN. Yes.

Mr. LUJAN. That would be a much easier task than the type of task that you performed?

Professor MARTIN. Yes. I'm just trying to think of other examples that might help put figures on this. At present the Royal Society in the United Kingdom is carrying out a study on the health of British science. They realized at the start there was no way that they could look at the whole of science and, therefore, they decided to choose two illustrative fields—one where it was felt amongst the scientists Britain was fairly strong; another where it was felt that Britain was not as strong as it might be.

They, then, have a staff of—what—about two full-time researchers plus input from the Fellows of the Royal Society, plus one of the officials from the Royal Society who looks after this work. So you're talking, again, about two or three man-years of effort going into this study to look at two fields.

Mr. LUJAN. Thank you.

Mr. FUQUA. Mr. Packard?

Mr. PACKARD. I have no questions, Mr. Chairman. Thank you.

Mr. FUQUA. Professor Martin, based on your form of analysis, have you arrived at any conclusions concerning what government policies tend to yield the best scientific results? That is very much part of the study we are doing.

Professor MARTIN. Yes. No, that's a much broader issue than we have attempted to look at in any of our work. We, as I tried to ex-

AVAILABLE

plain, have decided that the best way of gradually building up our activities and science policy—quantitative science policymaking—work in the UK is through relatively small-scale case studies, looking at individual fields. And on that basis, we feel that one can begin to get a feel for the strengths and weaknesses of different approaches, and, therefore, over the 7 years that we've been doing this work there's been a gradual and quite considerable evolution in the methods that we have been using.

It's only in the last year that we've begun to look at a broad issue, perhaps somewhat similar to the one you've mentioned, which is: How does British science stand at the present and what have been the trends over time, using, as I mentioned, these data compiled for the National Science Foundation by CHI Research? That's still in its very early stage. A lot of discussion and validation will now have to go on in order to provide the evidence that these numbers that we're beginning to get out of this have some significance. That process of validation, negotiation, and discussion could easily go on for a year or so before senior scientists and policymakers build up their confidence in these sorts of approaches.

But I think it is interesting that already there are one or two prominent people in the UK who are beginning to have such confidence in these methods. I'm thinking in particular of the chairman of the Advisory Board for the Research Council, who started, like most scientists, a few years back with a skepticism toward these methods, but he has been encouraging our work and other work in the field. He's provided us with two of the studies that we've been doing, one being this comparison of funding that we're working on at present.

And although obviously one would have to ask him about this, my impression is that he is gradually becoming a lot more confident in the numbers that come out of this and is beginning to see their relevance to the sorts of policymaking decisions which he is involved with as chairman of the Advisory Board of the Research Councils.

Mr. FUQUA. Your work has been, I gather, largely entered—or you related to it in your paper—on large laboratories.

Professor MARTIN. Yes.

Mr. FUQUA. How about evaluating colleges and universities, and particularly departments? The specific question I'm interested in is: we think that that is the resource whereby scientists are trained and is very important; that without the colleges and universities, then we would not have scientists working at Fermilab, CERN, or other places.

Professor MARTIN. Yes.

Mr. FUQUA. How do you take into account the training of future scientists not versus, but in addition to scientific papers and science that is produced there?

Professor MARTIN. OK. Well, first of all, you're absolutely right—it is important to look at university and college departments, and in fact that's, hopefully, the next project which we may be doing in the coming months.

As I explained, we decided, when we were beginning our work seven years ago, to focus on the big laboratories because the British science is so heavily concentrated on those.

But now, as I explained at the start, we have been diversifying and beginning to look at less capital-intensive areas of science, and there probably the best unit is, as you say, the university departments or maybe a research group within a university department.

So we hope in the near future to begin exploring the extent to which the indicators we've used on big science facilities can be applied to university groups.

I would suspect—and obviously we haven't done the work yet—that we will find they work relatively well for areas of more basic science, but for more applied areas, then the approach will be more similar to that we adopted in Norway, where the scientific literature-based indicators were less useful, and perhaps not useful at all in the case of mechanical engineering, and we had to rely on this combination of peer review—what did the researchers concerned think about the utility of their work—and customer review. In other words, what did the potential customers in industry think about it?

So one will need different indicators for different types of science.

The other point you raise related to training, which is extremely important, and I did not go into details here because I just wanted to focus on one study to show how it worked in practice. However, we have attempted, in at least one case, to look at the training function associated with university research, and that was, again, the very first study we did of radio astronomy, where we were interested in the people who had done radio astronomy as postgraduate students. They had done master's or they had done Ph.D.'s. Most of those are no longer radio astronomers. There's a limited job market for radio astronomers. Most of them have, therefore, moved on into a variety of positions.

And what we did was carry out a survey of, as far as possible, all those people who had done radio astronomy as postgraduate students from the time of the Second World War up to when we did the study in about 1978. We managed to get 70 percent of them to reply to a questionnaire, and in that questionnaire we asked three sets of questions.

One, what skills did you develop as a postgraduate student?

Two, what jobs have you done since then? And we were able to follow people through successive stages of their careers. As I say, some of them had graduated in 1945, some 25, 30 years ago.

And, three, in those jobs, which are the skills that you developed as a postgraduate have been most useful?

So from the results of that study, we were able to build up a picture of the flows of trained personnel from radio astronomy into a variety of jobs. A lot, for example, are now working in the telecommunications industry, which is, as you know, a growing and important area of industry.

And, secondly, we were able to see the sorts of skills which they developed as students which had been most useful in the jobs that they had moved into.

So I would see similar sorts of approaches for other areas perhaps generating the sorts of information related to this training function which you mentioned.

Mr. FUQUA. Anything further?

[No response.]

Professor Martin, thank you very much for being here this morning. It has been very interesting and you have a very unique approach to evaluation. We appreciate your input.

Professor MARTIN. Thank you very much.

[Whereupon, at 9:40 a.m., the task force recessed to reconvene at the call of the Chair.]

AVAILABLE

APPENDIX 1

SELECTED PAPERS BY MARTIN AND IRVINE

INTERNAL CRITERIA FOR SCIENTIFIC CHOICE: AN EVALUATION OF RESEARCH IN HIGH-ENERGY PHYSICS USING ELECTRON ACCELERATORS

By
BENJAMIN R. MARTIN AND JOHN IRVINE

Reprinted from
MINERVA
Vol. XIX, No. 3
Autumn 1981

MINERVA
59 St. Martin's Lane, London WC2N 4JS

AVAILABLE

Internal Criteria for Scientific Choice: An Evaluation of Research in High-Energy Physics Using Electron Accelerators

BENJAMIN R. MARTIN AND JOHN IRVINE

THE economic situation of scientific research is now very different from what it was in the early 1960s when Dr. Alvin Weinberg opened the debate on the criteria for scientific choice.[1] Annual rates of growth of 10 per cent. or more in the budget for science were then common in most Western countries, while today scientists face the prospect of no growth at all or even a decline. Some progress has also been made in developing techniques for the evaluation of the scientific performance of research groups. These two facts make it interesting to reconsider the question of scientific choice.

The Need for Evaluation in Large-scale Research

As recently as 50 years ago, basic scientific research was still a relatively inexpensive activity. Since the Second World War, however, the costs of research have soared, even allowing for inflation. During this period, science has flourished, with numerous new specialities starting, growing and maturing, many based on technological innovations originally created to meet industrial or military needs. Prominent among these new specialities are the "big sciences" such as high-energy physics, astronomy, and space research, where a single new facility now costs tens or even hundreds of millions of dollars. With this dramatic change in both the nature and scale of research activity, the need for procedures and criteria which will help in determining the choices necessary for the conduct of science policy has become all the greater. When the sums of money expended on basic scientific research were relatively small, making these choices was not such an urgent concern, except to the scientists themselves. However, as research has grown more costly, it has become all the more important that scientific resources should be used as effectively as possible. As Dr. Weinberg pointed out:

> It is only when science really does make serious demands on our society—when it becomes "Big Science"—that the question of choice really arises.[2]

Explicit and systematic policies for basic scientific research can be effectively constructed only on the basis of detailed information concerning the specific results achieved in different areas of research, and on the relative efficiency with which these results are achieved by different research groups. To facilitate the collection and appraisal of such information, scientists must be more willing to have the results of their research scrutinised by other scientists, especially those working in different specialities, as well as by the wider public. Scientific autonomy, in the sense of the freedom of scientists to decide how the resources for science should be distributed, must eventually be reconciled with considerations of efficiency when the expenditures involved exceed a certain critical size. For the "big sciences", that point has been reached. Dr. Weinberg, himself a physicist, recognises that:

---

[1] Weinberg, A. M., "Criteria for Scientific Choice", *Minerva*, I (Winter 1963), pp. 159–71. The main contributions in the debate are reprinted in Shils, E. (ed.), *Criteria for Scientific Development* (Cambridge, Mass.: MIT Press, 1968), esp. pp. 21-142.

[2] Weinberg, A. M., *op. cit.*, p. 171 and Shils, E. (ed.), *op. cit.*, p. 33.

> It is as much out of a prudent concern for their own survival, as for any loftier motive, that scientists must acquire the habit of scrutinising what they do from a broader point of view than has been their custom. To do less would cause a popular reaction which would damage mankind's most remarkable intellectual attainment—modern science—and the scientists who created it and must carry it forward.[3]

Such close scrutiny of scientific activity is essential not only to the construction of more effective science policies, but also to ensure that, in an era of growing demands for accountability, the public is provided with the necessary assurances that government funds expended on research have been employed efficiently in the past, and are likely to continue to be so employed in the future.

Four main questions need to be considered in any attempt to construct a national science policy: What should be the total expenditure on basic scientific research compared with other objects of public expenditure? How should this sum be distributed among the different disciplines, with their competing claims for funds? How much should be allocated to the different types of scientific institutions—university departments, regional or national laboratories, and international centres? What resources should be given to each research centre, group, or individual working within a discipline?

All these questions were to some extent discussed during the 1960s in the debate on criteria for scientific choice. To render the making of decisions on such questions more systematic, Dr. Weinberg proposed a number of criteria, distinguishing in particular between "internal" and "external" criteria:

> Internal criteria are generated within the scientific field itself and answer the question: How well is the science done? External criteria are generated outside the scientific field and answer the question: Why pursue this particular science?[4]

Dr. Weinberg's ideas subsequently attracted much attention in the scientific community and among the makers of science policy. However,

[3] Weinberg, A.M., *op. cit.*, p. 171 and Shils, E. (ed.), *op. cit.*, p. 33.
[4] Weinberg, A.M., *op. cit.*, p. 163 and Shils, E. (ed.), *op. cit.*, p. 25.

participants in the debate concentrated rather more on the external than the internal criteria; many were interested primarily in the level of expenditure on basic research, basing their arguments as to what that level should be either on the notion of an overhead charge in applied research, or on the idea of basic science as a part of "high civilisation".[5]

The situation facing science in the 1980s is very different, however, from that in the 1960s. The reallocation of governmental expenditures in recent years means that science must, at least in the near future, face the prospect of little or no growth at all.[6] Consequently, decisions concerning the distribution of the funds for science are beginning to assume greater importance than those concerned solely with the total expenditure on basic research.

The onset of a period of stable expenditure for science clearly makes the problem of determining choices within science more acute. If resources are to be made available for the support of promising new fields of science, this can only be done by diverting funds from old, declining specialities, or from research groups which, despite the "healthiness" of their field, have worked relatively unsuccessfully. While redistribution of this nature can be based in part on the use of external criteria, it must also take account of internal factors: we should ask which fields of science and which research groups are making or are likely to make the greater contributions to science.

However, any attempt to redistribute resources is likely to encounter many difficulties. The specialities and groups that have received substantial support in the past are likely to have become entrenched within the bodies that allocate funds for research. Because the representation of scientists on committees allocating funds is generally in proportion to prevailing levels of scientific support, "organisational inertia" may act against change in the distribution of resources. As a result, certain specialities or research groups may still receive relatively large sums despite a decline in their scientific achievements.

In some circumstances, it may well be necessary to cease supporting certain research activities altogether in order to release resources that can be employed more fruitfully in other areas of science. The desire to protect individual or institutional interests may, however, impede the redirection of resources towards promising new areas of science or towards more successful groups in fields already being supported. Consequently, there is a danger that the internal regulation of research by the judgement of peers will be unable to cope with the political pressures hindering the shift of resources towards more meritorious recipients. This danger is greater where research in a field is concentrated in a few centres and where the judgements as to the allocation of funds are made by panels of members drawn largely from the users of those few centres. In such a situation there are not enough "neutral peers". For this reason, a supplementary or alternative mode of assessment of scientific achievements should be considered. Such an assessment can be based in part on tools developed over the past few years within the "science of science".

---

[5] See, for example, Toulmin, S., "The Complexity of Scientific Choice II: Culture, Overheads or Tertiary Industry?" *Minerva*, IV (Winter 1966), pp. 155–169 and Shils, E. (ed.), *op. cit.*, pp. 119–133.

[6] For example, the best that can be expected for British basic science over the first half of this decade is that expenditure will grow in real terms by 1 per cent. per annum, and it might even decline over the early years. See *The Government's Expenditure Plans, 1980–81 to 1983–84*, Cmnd. 7841 (London: H.M. Stationery Office, 1980).

The Quantitative Assessment of Scientific Achievement

Many of the limitations of measures such as the numbers of research publications, or the numbers of references to those publications, have been ignored in their early, often indiscriminate, use in scientific assessment. We have tried to devise a means of assessing scientific achievement which is more realistic. It is based on a number of converging "partial indicators".

We have applied this method to evaluate the achievements of a group of high-energy physicists who used the major British electron accelerator, NINA, operated by the Daresbury Laboratory. This accelerator was closed in 1977, a decision which met with considerable opposition from many British high-energy physicists. One principal argument advanced by the opponents of the decision was the lack of systematic scientific information put forward to justify the decision. The arguments over the future of the accelerator emphasized its costliness rather than being based on a detailed evaluation of the contributions to scientific knowledge made by the physicists using the accelerator.

Since the end of the Second World War, high-energy physics has received by far the largest part of British expenditure on basic science. From the late 1950s to the mid-1970s, this field of physics annually accounted for about 40 per cent. or more of the British basic science budget. In the financial year 1963-64, for example, Britain spent nearly twice as much on high-energy physics as on all other fields of basic natural science.[7] For much of the period up to the mid-1970s, about half of the appropriation for high-energy physics was used to pay Britain's contribution to the European Organization for Nuclear Research (CERN); the remainder was largely accounted for by the two national high-energy physics centres at the Rutherford and the Daresbury Laboratories, with a relatively small amount for direct grants to individual scientists at

---

[7] In 1963-64 total expenditure on the National Institute for Research in Nuclear Science (NIRNS)—the body responsible for the operation of Daresbury Laboratory as well as the other national accelerator centre at Rutherford Laboratory—amounted to £8·4 million. To this must be added the British contribution of £2 million to CERN and a major part of the £1 million in research grants allocated to universities for nuclear physics research by the Department of Scientific and Industrial Research (DSIR), making a total of about £11 million. In contrast, DSIR gave grants of £2·7 million to the universities to support all other fields of scientific research, and £2 million for awards for postgraduate training. In addition, DSIR was responsible for £1·4 million expenditure on British contributions to international organisations other than CERN, yielding a total of £6·1 million which is only slightly more than half the £11·4 million spent on high-energy physics.

AVAILABLE

universities. This high degree of concentration of resources on three centres within one field of physics greatly affected the scale of activity throughout the rest of British basic science.

The Daresbury Laboratory was established in 1963 to operate a 4 GeV electron accelerator, mainly serving university physicists in the north of the United Kingdom.[8] Financial support for Daresbury amounted on average to some 8 or 9 per cent. of the annual budget of the Science Research Council,[9] and, by the time NINA was closed in 1977, a total of over £40 million had been spent on high-energy physics at the laboratory.[10] In assessing the correctness of the decision to discontinue the operation of NINA a number of questions are pertinent. The main question is: exactly what returns did this expenditure of £40 million bring? More specifically: what contributions to scientific knowledge did NINA make? The construction of NINA had been authorised in order to enable British scientists to undertake "the investigation of quite new fundamental problems in the physics of elementary particles".[11] To what extent did it contribute to the solution of "fundamental problems"?

The Assessment of Research Groups

Evaluations of the achievements of scientists engaged in basic research have generally been concentrated on the individual scientists or the speciality as a whole, rather than the research group. This is not always a helpful approach since most scientific resources tend to be allocated directly to research groups rather than to individuals or to specialities.[12]

The products of scientific research can take a variety of forms such as contributions to knowledge in the discipline concerned or in neighbouring disciplines, or the training of scientists; they can also take the form of contributions to industry in terms of technological knowledge and skilled persons. Because of the diverse and somewhat intangible nature of its products, no comprehensive and unitary quantification of the achievements of basic science is possible. In view of this, assessment of the performance of a scientific group must be based on a comparison of the achievements of one group with those of others utilising broadly similar amounts of money and working in the same scientific field. We can, of course, legitimately compare only "like" with "like".[13] It would not be reasonable, for example, to compare the scientific achievements, in terms of numbers of discoveries or costs per paper, of a group of high-energy physicists using an accelerator like NINA with those of a group of astronomers using a large telescope.

---

[8] The research of high-energy physicists at southern British universities was conducted mainly at Rutherford Laboratory or at CERN.

[9] During 1965–66, the first year of the Science Research Council's existence, the figure was 8·7 per cent. and in 1976–77, the last year of operation of the Daresbury accelerator, it was 8 per cent., although by then not all of the expenditure was for high-energy physics research.

[10] Up to the closure of NINA in 1977, a total of £61 million was spent on the facilities at Daresbury. However, £9·6 million of this was used to build the Nuclear Structure Facility, £3·2 million went to support of synchrotron radiation work, and £3 to 4 million was spent on supporting experiments at CERN.

[11] *Fifth Annual Report of the National Institute for Research in Nuclear Science, 1961–62* (London: H.M. Stationery Office, 1962), p. 14.

[12] See Irvine, J. and Martin, B.R., "A Methodology for Assessing the Scientific Performance of Research Groups", in *Proceedings of the Conference On Evaluation of Science and Technology: Theory and Practice, Dubrovnik, Yugoslavia, July 1980*, published in *Scientia Yugoslavica*, 6 (April 1980), pp. 83–95.

Since no two groups produce identical kinds of results, the task of selecting exactly matched groups for comparison might appear to be a virtual impossibility. Even two groups working in the same field and producing fairly similar results may have very different educational and technological effects. In high-energy physics, however, a certain degree of simplification is justified. High-energy physics is, above all, a basic science, the primary purpose of which is to increase scientific knowledge, rather than create educational, technological, or cultural benefits. In arriving at decisions on the distribution of resources within basic science, the main criterion must therefore be that of how great a contribution is likely to be made to scientific knowledge. This was Dr. Weinberg's fundamental internal criterion.

In the case of British high-energy physics, there have undoubtedly been external benefits in the form of technological "spin-off" and trained research workers for industry,[14] but these were not the primary reason for supporting research in this field.[15] Important external benefits can be no more than supplementary reasons to support scientific activities that are important according to internal criteria.

Dr. Weinberg suggested two closely related criteria for scientific choice: Is the field ripe for exploitation? And are the scientists in the field really competent?[16] He adds that, in practice, the second of these is normally the main one used in assessing applications for research grants. The question implied by the second criterion can best be answered by ascertaining how much the work of the group of scientists in question has contributed to the advance of knowledge. Past achievement, of course, is not the only clue to future performance, but it is undeniably one of the most important. If other factors such as the "ripeness" of the proposed research topic or its potential for scientific exploitation are equal, or indeterminate, a new scientific project clearly stands more chance of being carried out successfully by a group with a record of major contributions to the advancement of knowledge over the recent past than by one with an undistinguished record.

---

[13] See Toulmin, S., "The Complexity of Scientific Choice: A Stocktaking", *Minerva*, II (Spring 1964), p. 354 and Shils, E. (ed.), *op. cit.*, pp. 63–79.

[14] Some of the more important economic benefits to firms are pointed to in Schmied, H., "A Study of Economic Utility from CERN Contracts", *IEEE Transactions on Engineering Management*, EM-XXIV (1977), pp. 125–138.

[15] However, the probability of technological by-products has been advanced as a reason for supporting very high-cost basic science. See Richter, B., "Statement before the House Committee on Science and Technology, December 10, 1979" (Stanford Linear Accelerator Center, Stanford University) mimeograph, for a review of some of the more important developments. Even in this case, such "spin-off" would hardly constitute a justification for supporting mediocre research.

[16] Weinberg, A. M., *op. cit.*, p. 163 and Shils, E. (ed.), *op. cit.*, p. 25.

AVAILABLE

Dr. Weinberg's first internal criterion should perhaps be reformulated in two parts: What is the state of the field itself—is it ripe for exploitation? And how well placed is the applicant-group in comparison with its probable competitors—does it possess some advantage enabling it to exploit the field and to achieve significant scientific results better than they are likely to do? In deciding whether a field is "ripe" for scientific advance, it is difficult to see how one can avoid depending on the opinions of scientists in the field itself. These would comprise, on the one hand, the proponents of the research project who, presumably, would not have proposed the project unless they were convinced of the field's ripeness for advancement, and, on the other, those scientists who are engaged in rival activities within the field, and who are to some extent in competition with the project's proponents. Hence, the opinions of neither body of scientists may be unqualifiedly neutral and objective. Even scientists in neighbouring fields, who constitute a third group of possible referees, may also make recommendations which are at least partly influenced by their desire to assure money for their own research.

The ability of the applicant-group to exploit a field ahead of its competitors is probably more amenable to systematic evaluation; the previous achievements of the group can be compared with those of its rivals. If competitors with similar research facilities have a better record over the recent past, then the group applying for support cannot generally be regarded as being in the best position to exploit the scientific potentialities of that field. The only exception to this is where a highly innovative research instrument is proposed which would place an entirely new group, or a group that had not done well in the past, in a unique position within its field. The procedure which we propose cannot determine the "ripeness" of a field for "exploitation"; at present only the judgement of qualified scientists can offer guidance here. Once such a decision is made, however, officials charged with judging applications for grants should pay attention to the relative merits of the applicants as ascertained by a scrutiny of their respective previous achievements.

Indicators of Scientific Performance

Since most research results are published in journals or conference proceedings, one obvious way to assess scientific performance is by the number of papers published by a scientist or group.[17] But all papers are not equal contributions to science. Data on the "numbers of papers" can be combined with data on the financial resources needed to produce these papers to yield such figures as the "cost per publication" and the "publications per research worker per year". This does not necessarily tell us anything about a group's relative contribution to scientific knowledge, however. Practices in publication vary considerably between research groups; differences in the pressure to publish in order to enhance the prospects of promotion or of future support may result in differences in the rate of publication. Furthermore, one group may conduct several short research projects, yielding many papers, while another engages in a single long-term programme that results in few papers. Consequently, while numbers of publications may permit a reasonable estimate of scientific productivity, they are only partial indicators of scientific achievement.[18]

---

[17] For a summary of some investigations along these lines, see Gilbert, G. N. and Woolgar, S., "The Quantitative Study of Science: An Examination of the Literature", *Science Studies*, 4 (July 1974), pp. 279–294.

What other indicators can be employed apart from those based on the number of publications? One is the number of times the published work of a research group is cited by other scientists. Citations have been used as a measure of the "quality" of scientific research in much previous work on this problem. There are, however, many difficulties in this procedure. The *Science Citation Index*, which is generally used in these studies, does not provide a sufficiently complete record of citations. For example, citations are credited only to the first-named author of papers with more than one author; there are variations in authors' names (for example, in the use of one rather than two initials); there may be several authors with identical names; and finally there are numerous errors arising from the failure of authors to give the correct bibliographical references. The first of the difficulties is less severe when we deal with a group of scientists, since all citations are credited to the group and not to individuals, while many of the other problems can be largely overcome by careful checking.[19]

Despite this, serious difficulties in the use of citations remain. For one thing, certain important papers initially go unrecognised in the scientific community and therefore receive few citations. Second, important papers often present ideas which are so quickly and completely incorporated into the body of scientific knowledge that the original papers eventually cease to be cited. In addition, some papers may be referred to critically—because they are viewed as "mistaken"—while others may be cited purely because the prominence of their authors is thought to add authority to an argument.[20] However, if the number of citations to a paper is regarded as an indication of its impact on the advancement of scientific knowledge rather than its intrinsic scientific quality or importance, then many of the problems of citation-analysis become less severe.

Here, we define the "quality" of a research paper in terms of the absence of obvious "error", the aesthetic merit of the mathematical formulations, the originality of the conclusions, and so on. The "import-

---

[18] See Martin, B. R. and Irvine, J., "Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy", *Research Policy* (forthcoming).

[19] This is the reason why we did not use the computer tapes on citations available from the Institute for Scientific Information, which publishes the *Science Citation Index*.

[20] See Smith, R. and Fiedler, F. E., "The Measurement of Scholarly Work: A Critical Review of the Literature", *Educational Record* (Summer 1971), pp. 225–232. These and other problems in citation-analysis are discussed at length in Martin, B. R. and Irvine, J., *op. cit.*

AVAILABLE

ance" and "impact" of a research paper, in contrast, refer to the relations between the scientific results reported and current work in associated areas of research. The "importance" of a publication describes its potential influence on neighbouring research activities—that is, the influence it would have on the advancement of scientific knowledge if there were perfect communication in science. However, there are imperfections in the system of scientific communications; for example, a paper by a young scientist at a provincial university, published in a journal in a language other than English and having a small circulation, may go unnoticed by scientists regardless of its merits. Therefore, the "importance" of a paper is generally not synonymous with its "impact". It is the "impact" of a paper which represents its actual influence at a particular time on surrounding research activities and hence on the advancement of knowledge,[21] and it is this impact, rather than its quality or importance, that is reflected in the number of citations earned by a particular paper.

While the impact of the published work of a group or research unit will depend to a large extent on the work's importance, it is also influenced by such factors as the status of the institution in which the group is working, and the prestige, language and circulation of the journals in which its results are published. The number of citations received by a publication is an indicator of its impact. However, it is by no means a perfect measure of impact—the frequency of citation may vary with type of paper and with speciality, which is why citations can only be used to compare matched groups producing similar types of papers within the same speciality. Since it is the impact rather than the importance of the publication which reflects its contribution to scientific knowledge, we may regard the number of citations as a partial indicator of the scientific achievements of a research group.

The method of assessing contributions to scientific knowledge generally favoured by scientists is "assessment by peers". However, again this is not a wholly adequate measure of scientific achievement. Any assessment made by scientists of the work of their peers is subject to influence by social and political considerations within the scientific community. Few scientists are unaware of the fact that such assessments may have repercussions on the future support of both their own group and their competitors. Furthermore, individual scientists can only assess the scientific contributions of others in relation to their own cognitive locations,[22] and assessments may differ. Finally, scientists tend to conform to conventionally accepted assessments; they may feel it necessary to praise publicly the work of eminent scientists, even if privately they hold that work in low esteem; they might also be ignorant of the work of a particular research group and might therefore make assessments more on the general reputation of that group within the scientific community than on the particular research which is allegedly being evaluated.

---

[21] One of Dr. Weinberg's internal criteria: "How well is the science done?" should therefore perhaps be reformulated as "How much impact on the advance of knowledge does the research have?"

[22] See Martin, B. R., "Cognitive and Social Locations: Their Role in the Processes of Discovery and Evaluation within Science", (Department of Liberal Studies in Science, University of Manchester: 1977), mimeograph.

To reduce the possible distortions in the use of assessments by peers, we interviewed a large, representative sample of scientists, assuring them of confidentiality. When a divergence was suspected between opinions expressed and actual views held, the scientists being interviewed could be pressed further. Even so, it is clear that assessment by peers, like citations and publications, gives no more than a partial indication of the scientific contributions of research groups.

In assessing the scientific achievements of the group working with NINA relative to those of groups of scientists working with other accelerators, we have therefore employed indicators based on publications, citations and peer-evaluation. By counting the total number of experimental papers produced annually by the group using each accelerator, we were able to calculate the average cost per publication and the publication rate for each research centre. We considered both the average number of citations per paper for the papers produced by each group (the "citation-rate") and the number of very highly cited papers. The former provides an indicator of the average influence of all the papers produced by the users of each particular accelerator, while the latter is indicative of the number of major "discoveries". Both indicators are essential, since scientific progress is made not only in the form of gradual change—to which all publications may be said to contribute—but also through the occasional, revolutionary transformations resulting from only a small number of major papers. A group with relatively low rates of publication and citation may still have made a substantial contribution to scientific knowledge if it has been responsible for one or two major discoveries, reported in papers which have been cited with especially high frequency. Assessment by peers complements these two measures, providing some sort of weighted average indicator for both the great mass of relatively minor contributions made by the users of an accelerator and the small number of major advances.

NINA and its Closest Competitors [23]

We have attempted to compare the contributions to scientific knowledge made by the users of the Daresbury synchrotron (NINA) with those associated with the other major electron accelerators in Europe and in the United States. There were originally three electron synchrotrons in operation which were sufficiently similar to NINA in energy-range to

---

[23] This assessment forms part of a larger project which attempted to assess the performance of five major centres supported by the British Science Research Council. Beside Daresbury and Rutherford Laboratories, we examined the two radio astronomy observatories at Cambridge and Jodrell Bank, and the national optical astronomy facilities at the Royal Greenwich Observatory. The results of the study of the radio-astronomy observatories appear in Martin, B. R. and Irvine, J., *op. cit.*, and the results of the optical astronomy observatory in Irvine, J. and Martin, B. R., "Assessing Basic Research: The Case of the Isaac Newton Telescope", *Social Studies of Science*, XIII (February 1983), pp. 49-86.

AVAILABLE

make possible direct comparisons of results: the Cambridge Electron Accelerator (CEA) operated by Harvard University and Massachusetts Institute of Technology; the Deutsches Elektronen Synchrotron (DESY) at Hamburg; and a Russian machine at Yerevan in Armenia, although we have very little data on this, apart from some judgements by individual scientists. The energies of the four accelerators ranged between approximately 4 and 6 GeV, and all were completed in the 1960s. The only other major electron accelerator built during this period was the 20 GeV Stanford Linear Accelerator (SLAC), and this was also included in our study (Figure I).

FIGURE I

The World's Main Electron Accelerators

Centre	Details of Accelerator(s)	Completion date
Cambridge Electron Accelerator (CEA)	6 GeV Electron Synchrotron	1962
	BYPASS—3·5 GeV Electron-Positron Storage Ring	1972
Daresbury Laboratory	NINA—4·5 GeV Electron Synchrotron	1966
Deutsches Elektronen Synchrotron, Hamburg (DESY)	6 GeV Electron Synchrotron[a]	1964
	DORIS—3 GeV Electron-Positron Storage Ring[b]	1973
	PETRA—19 GeV Electron-Positron Storage Ring	1978
Stanford Linear Accelerator Center (SLAC)	20 GeV Linear Electron Accelerator	1966
	2 SPEAR—2·5 GeV Electron-Positron Storage Ring	1972
	PEP—18 GeV Electron-Positron Storage Ring	1980
Yerevan	6 GeV Electron Synchrotron	1967

[a] Increased to 7·5 GeV in 1968.
[b] Increased to 5 GeV in 1976.

Source: Annual reports from the centres.

Measures of the Scale of Research at Each Accelerator

Before examining the various indicators of scientific achievement, it is first necessary to consider the scale of research activity carried on at each accelerator; in particular, the number of scientists doing research at each

TABLE I

Numbers of High-Energy Physicists using the Four Centres[a]
1966-1978

		CEA (Cambridge)	DESY (Hamburg)	NINA (Daresbury)	SLAC (Stanford)
	Central research staff	19	~60	~15[c]	~60[c]
1966	External experimental users	54	~25	~30[c]	~40[c]
	Research students	56	~15	~20[c]	~20[c]
	Total	129	~100	~65[c]	~120[c]
	Central research staff	19	~70	27	~80
1970	External experimental users	23	~50	36	~70
	Research students	27	~30	26	~50
	Total	69	~150	89	~200
	Central research staff	10	~90	19	~90
1974	External experimental users	12	~70	28	~160
	Research students	8	~30	13	~50
	Total	30[b]	~190	60	~300
	Central research staff	—	~100	4[d]	~100
1978	External experimental users	—	~130	19[d]	~200
	Research students	—	~70	12[d]	~60
	Total	—	~300	35[d]	~360

[a] This excludes synchrotron-radiation research workers, but includes high-energy theoretical physicists resident at each of the centres, since their research is generally so closely related to the experimental work as to constitute an integral part of it. The errors in these figures should be no greater than 10% or 15% at most.

[b] Figure for 1972, the last year in which the Cambridge centre was operating fully.

[c] Figures for 1967, the first complete year of operation of the accelerators.

[d] Figures for 1976-77, which was the last year in which NINA was operating.

SOURCE: Annual reports from the centres and private correspondence.

of the accelerators has varied considerably both among the centres and over time (Table I).

Almost from the time the Stanford Linear Accelerator began operation, it has supported the largest research group of the four. By 1970, it was providing experimental facilities for over twice as many physicists as the Daresbury synchrotron, and about 30 per cent. more than the German synchrotron at Hamburg. The group using the Cambridge accelerator was by this time decreasing in size, largely as a result of a major explosion in the late 1960s which seriously restricted its research programme; by 1970, it provided experimental facilities for only approximately three fourths of the number using NINA at Daresbury. Up to 1974, the ratio of the numbers of physicists at Hamburg and at Stanford stayed at about 2:3; both centres had grown rapidly, so that by this time the number of research workers using the Hamburg centre was three times that using Daresbury. In the meantime, the centre at Cambridge had closed because of the American policy of concentration of high-energy physics activity in three major centres. Between 1974 and 1978, the number of users of the Hamburg centre continued to grow, expanding by more than 50 per cent. to approach the size of the group using the Stanford accelerator. In contrast, NINA served a steadily decreasing number of physicists until it was closed in 1977.

In estimating the scale of research activity associated with an accelerator, it is necessary to bear in mind that not all users were able to spend all their time on research; some of them had other tasks as well, not least the teaching of undergraduates.[24] The external users in particular had to spend substantial fractions of their time on teaching (Table II).

TABLE II

Numbers of "Effective" High-Energy Physicists using Each Accelerator 1966–1978

	CEA	DESY	NINA	SLAC
Percentage of time spent in teaching:[a]				
Central research staff	5%	5%	5%	5%
External experimental users	25%	30%	40%	25%
Total number of "effective" research workers after allowing for teaching:				
1966	115	~90	52[c]	~105[c]
1970	62	~130	73	~180
1974	27[b]	~165	48	~255
1978	—	~255	27[d]	~305

[a] In calculating the "effective" number of physicists, we have assumed that the percentages reported in 1979–80 did not change appreciably over the preceding 10–15 years.

[b] 1972. See Table I, note b.

[c] 1967. See Table I, note c.

[d] 1976–77. See Table I, note d.

SOURCE: Annual reports from the centres and private correspondence.

Another major factor determining the scale of research activity is obviously the annual operating costs. Unfortunately, variations in accounting procedures between the centres make it difficult to obtain fully comparable figures (Table III).[25] However, it is clear that the Cambridge Electron Accelerator always had the lowest running costs. In 1970, the

---

[24] *I.e.*, students reading for first degrees: B.Sc. in Britain, B.A. or B.S. in the United States and *Diplom* in Germany. However, no allowance has been made here for time spent on administrative activities, nor on the construction and maintenance of equipment, the justification being that such duties must be regarded as an integral part of modern, large-scale research.

[25] For example, it is not clear whether the costs of detectors for accelerators (these typically now cost millions of pounds) should be included in the capital costs of the facility or in its recurrent costs. Certain centres, too, admit to having reallocated resources to capital spending from operating costs when the supporting bodies have refused to finance new projects.

TABLE III

Annual Operating Costs[a] *(in pounds sterling)*[b]

	CEA	DESY	NINA[c]	SLAC
Annual operating costs (in millions) in:				
1966	£1·3	£3·4	£2·7[f]	£6·4[f]
1970	£1·3	£4·9	£3·9	£10·0
1974	£0·9[e]	£10·9[d]	£3·2	£10·2
1978	—	£20·3[d]	£1·[illegible][g]	£14·5
Annual cost per physicist (including students) (in thousands) in:				
1966	£10	~£35	£25[f]	~£55[f]
1970	£20	~£35	£35	~£50
1974	£30[e]	~£55	£55	~£35
1978	—	~£70	£50[g]	~£55

[a] Operating costs are exclusive of salaries of visiting physicists, which are paid by their home institutions.

[b] These figures have been calculated using the average annual exchange rates quoted in the *National Institute Economic Review* (No. 86, November 1978, Table 25).

[c] See Table I, note b.

[d] These figures have been calculated by adjusting the actual total operating cost to allow for the effort put into synchrotron radiation work (approximately DM 1·2 million per annum).

[e] These figures have been calculated by adjusting operating costs to take account of synchrotron radiation and nuclear structure work.

[f] Figure for the financial year 1967–68, the first complete year of operation for the accelerator.

[g] Figure for 1976–77, the last year of NINA's operation.

SOURCE: Annual reports from the centres, interviews and private correspondence.

operating costs for the Daresbury synchrotron were approximately three times those of the Cambridge machine. The next most expensive was the German machine. By 1974 its operating costs had grown so rapidly that they were similar to those of the Stanford Linear Accelerator, both being just over three times larger than those of NINA. During the most recent four-year period, the costs of the Hamburg and the Stanford establishments have both almost doubled as new storage-ring facilities were built and brought into operation (Figure I).

As with operating costs, capital costs are not strictly comparable, again because of differences in accounting procedures among centres; there were, for example, differences over whether or not labour costs internal to the centre, or the costs of certain buildings, should be included as expenditure on capital equipment. Despite this, it can be seen that, of the four accelerators, the Cambridge synchrotron was by far the least expensive; the Daresbury accelerator cost over twice as much, and the Hamburg synchrotron three times as much, while the cost of the Stanford Linear Accelerator was six times that of the Hamburg synchrotron. The Stanford figure is particularly high because it contains a substantial amount for "salaries". We therefore calculated figures for the total expenditure, *i.e.*,

BEST COPY AVAILABLE

TABLE IV

Energy and Capital Costs of Original Accelerators[a]

	CEA	DESY	NINA	SLAC
Energy	6 GeV	6 GeV	4 GeV	20 GeV
Quoted capital cost[b] (in millions of pounds sterling)	£1·9	£6·7	£4·4	£40
Date of first experimental publication	1963	1965	1969	1967
Total expenditure up to year of first publication (in millions of pounds sterling)	£5·5	£13·7	£14.8	£66

[a] See Table III, note b.

[b] As far as possible, this figure is intended to represent the capital cost of the accelerator and the accelerator buildings, but to exclude the cost of other developments on the site (central offices, computers, etc.). However, the Stanford figure includes a much higher element for the salaries of staff-members employed in the design and construction work than do the figures for the other centres. It should be noted that these figures have not been adjusted for inflation; this may partly explain the slightly lower capital cost of the Cambridge accelerator, which was the first of the electron synchrotrons.

SOURCE: Annual reports from the centres and private correspondence.

all capital and recurrent costs, up to the year in which the first experimental results were published (Table IV). The costs for the Hamburg synchrotron and NINA were very similar, while the Stanford accelerator was considerably more expensive. Its costs were four times greater; this is consistent with the ratio of the energies of the machines.

One final factor that influences the scale and productivity of research conducted with an accelerator is the numbers of supporting staff serving the high-energy physicists (Table V). In 1966, the users of the Hamburg centre had about 70 per cent. more supporting staff than Daresbury; at each, the number grew by about 30 per cent. over the next four years. During the 1970s, however, the number of supporting staff for users of the Daresbury accelerator decreased as more effort was put into research on synchrotron radiation and nuclear structure, while for Hamburg the number continued to rise. Once the accelerators were in operation, the ratio of supporting staff to physicists for Hamburg, NINA and Stanford in 1970 was fairly constant at 6:1 although for Hamburg and Stanford it dropped to about 4:1 in 1978; for NINA it stayed at about 6:1.

If we consider all these measures together, they suggest that the scale of research activity at Stanford was initially considerably greater than at Hamburg, which was the largest of the other three centres. By the time of its first publication, the scale of research at the Daresbury synchrotron had overtaken the Cambridge facility in size, but it was never more than about 60 per cent. of that at the Hamburg centre. During the 1970s, the German centre grew rapidly, with the construction first of a 3 GeV electron-positron storage ring and then a higher energy (19 GeV) storage ring (Figure 1), so that by 1978 its level of activity was only perhaps 20

TABLE V

Numbers of Supporting Staff-members[a]

Numbers of supporting staff-members		CEA	DESY	NINA	SLAC
1966	Central	180	690	~380[d]	1200[d]
	User institutions	80	~40	~40[d]	~80[d]
	Total	260	~730	~420[d]	1280[d]
1970	Central	170	830	502	1200
	User institutions	35	~80	48	~140
	Total	205	~910	550	~1340
1974	Central	98	940[c]	345	1000
	User institutions	~30	~110	44	~320
	Total	~128[b]	~1050	389	~1320
1978	Central	—	950[c]	184[e]	1100
	User institutions	—	200	29[e]	~400
	Total	—	~1150	213[e]	~1500
Ratio of supporting staff to physicists:					
	1966	2·0	7·3	6·5[d]	10·7[d]
	1970	1·5	6·1	6·2	6·7
	1974	4·3[b]	5·5	6·5	4·4
	1978	—	3·8	6·1[e]	4·2

[a] These include not only supporting staff employed at each centre, but also those at the home institutions of the visiting physicists.

[b] See Table I, note b.

[c] These are the figures obtained after subtracting those staff engaged in synchrotron-radiation work.

[d] See Table I, note c.

[e] See Table I, note d.

SOURCE: Annual reports from the centres and private correspondence

per cent. smaller than that of Stanford, which had built a storage ring and was in the course of constructing another machine almost identical to the German 19 GeV storage ring.

Publications

Having obtained some idea of the respective scales of research activity at each accelerator, we can now consider the first of the indicators of scientific achievement—the number of experimental publications arising from the use of each accelerator (Table VI). Over the first ten years of its life, the Hamburg synchrotron appears to have been relatively productive, research carried out there resulting in 50 per cent. more papers than was the case for the Cambridge Electron Accelerator (13 per annum compared

AVAILABLE

TABLE VI

Numbers of Published, Experimental High-Energy Physics Papers from Electron Accelerators (excluding Work on Storage Rings)

Year	CEA	DESY	NINA	SLAC
1963	2	—	—	—
1964	6	—	—	—
1965	6	1	—	—
1966	10	10	—	—
1967	19	19	—	6
1968	12	19	—	13
1969	6	6	2	18
1970	7	16	6	13
1971	6	18	8	17
1972	2	15	10	18
1973	3	16	9	17
1974	—	9	4	23
1975	—	8	5	30
1976	—	10	6	30
1977	—	5	5	19
1978	—	12	4	40
1979	—	3	7	18
Total	79	167	66	262
Average no. per year over first ten years	8	13	6	19

SOURCE: Complete lists of publications supplied by the Daresbury, Hamburg and Stanford centres. Cambridge publications compiled from "Source Index" of *Science Citation Index*.

with 8 per annum at Cambridge). The production of papers at the Daresbury accelerator, like that of Cambridge and Hamburg, also reached its peak in the fourth year, before diminishing quite rapidly, but the average number of papers produced each year over its first decade of operation was only half that of Hamburg and three quarters of that of Cambridge. The linear accelerator at Stanford, having a greater number of users, produced far more publications than the accelerators at the other three centres; moreover, its rate of publication has actually continued to increase, more or less up to the present time as new uses have been found for the facility. Over its first ten years of operation, the research workers at this accelerator produced on average about 50 per cent. more papers annually than those using the synchrotron at Hamburg, two and a half times as many as those using the Cambridge accelerator, and three times as many as those at the Daresbury facility.

Cambridge and Hamburg initially had the highest productivity per

TABLE VII

Average Numbers of Publications in Relation to Numbers of Users and Financial Resources[a]

	CEA	DESY	NINA	SLAC
No. of papers per "effective" physicist[b] in:				
1966	0·13	0·18	0·01[d,e]	0·10[d]
1970	0·08	0·13	0·11	0·08
1974	0·15[c]	0·08	0·10	0·13
1978	—	0·11	0·20[f]	0·12
No. of papers per £m operating costs[b] in:				
1966	11·3	4·7	0·4[d,e]	1·9[d]
1970	3·8	3·3	2·4	1·6
1974	4·4[c]	1·2	1·6	3·8
1978	—	1·4	2·9[f]	2·2

[a] These figures should be accurate to between 10 and 15 per cent.

[b] Because the number of papers published in any one year is fairly small, the random fluctuations from year to year are relatively large. We have attempted to "smooth out" these fluctuations by averaging over a three-year period, beginning with the year in question. Thus the 1970 figures are calculated by averaging the numbers of publications in 1970, 1971 and 1972, and dividing by the number of "effective" physicists in 1970 (Table II). This procedure also allows for some of the lapse of time between carrying out an experiment and the publication of the results. Because the number of publications in 1980 were unknown at the time of writing, the 1978 figures are based on the average number of papers produced in 1978 and 1979.

[c] See Table I, note b.

[d] See Table I, note c.

[e] This figure is low because, although NINA was operating in 1967, the first publications did not appear until 1969. The average annual rate of publication for the period 1967 to 1969 is therefore low.

[f] See Table I, note d.

SOURCE: Complete lists of publications supplied by the Daresbury, Hamburg and Stanford centres. Cambridge publications compiled from "Source Index" of *Science Citation Index*.

"full-time" physicist (Table VII). However, at Hamburg, the productivity decreased as experiments, particularly with the storage rings, became more expensive and more labour-intensive. In contrast, the productivity of Cambridge, and especially NINA, increased at the end of their working lives as research carried out in previous years was written up and published, while the numbers of physicists still using the accelerators and the operating costs were both falling rapidly.

As we argued earlier, however, it is not sufficient to consider solely the numbers of papers yielded by each accelerator. It is, for example, quite conceivable that each of the papers from one accelerator represented, on average, a more significant contribution to scientific knowledge than those from the others. Until we have considered the impact of those papers, we cannot reach any conclusions about the relative contributions to scientific progress made by the accelerators.

Assessment of Impact

Because of its earlier beginning, the published work of the Cambridge centre was obviously the first to make an impact (Table VIII). By 1968, however, it had been overtaken by the Hamburg synchrotron, by the Stanford Linear Accelerator in 1970, and finally by NINA in 1976—four years after the Cambridge centre had closed down. (In 1978, however, papers originating from the Cambridge accelerator were more frequently cited than those from NINA.) It was not long before the Stanford centre overtook Hamburg, generally receiving between 50 per cent. and 100 per cent. more citations each year than the German accelerator, which, in turn, obtained between three and five times as many citations as work carried out at Daresbury. When the storage-ring work is included, these differences are further accentuated, with Stanford receiving four times as many citations as Hamburg in 1976 and over twice as many in 1978 (Table IX).

TABLE VIII

Citations of All Previous Electron-Accelerator Work (excluding Work on Storage Rings)

Citations in:	CEA	DESY	NINA	SLAC
1964	28	—	—	—
1966	139	15	—	—
1968	219	303	—	88
1970	160	306	19	405
1972	149	358	72	504
1974	103	324	78	506
1976	50	216	76	454
1978	76	305	57	585
Average no. per year over first ten years	140	260	60	390

SOURCE: Manual analysis of data from *Science Citation Index*.

Whereas the total number of citations is to a large extent determined by the amount of research activity at each centre, the average number of citations per paper takes account of this factor, so that a small but relatively successful centre, even if it produces fewer papers, should still be able to achieve a relatively large number of citations per paper. The papers produced at Stanford none the less appear to have had on average the greatest impact (Table X). In contrast, NINA was comparatively late on the scene. By the time of its first publications, many of the electron-synchrotron experiments likely to have a significant impact on scientific knowledge had already been carried out elsewhere. Moreover, even the

TABLE IX

Citations to All Previous Experimental Work Including Work on Storage Rings

Citations in:	CEA	DESY	NINA	SLAC
1964	28	—	—	—
1966	139	15	—	—
1968	219	303	—	88
1970	160	306	19	405
1972	149	358	72	504
1974	197	324	78	512
1976	67	295	76	1282
1978	81	644	57	1463

SOURCE: Manual analysis of data from *Science Citation Index*.

TABLE X

Average Citations per Publication for Work Published in last Four Years (excluding Work on Storage Rings)

Citations per publication in:	CEA	DESY	NINA	SLAC
1966	5·8	—	—	—
1968	4·1	6·2	—	—
1970	3·0	4·6	—	8·1
1972	4·0	4·4	2·8	6·5
1974	3·9	3·2	2·3	4·9
1976	—	2·0	1·9	3·0
1978	—	2·5	0·9	2·7

SOURCE: Manual analysis of data from *Science Citation Index*.

best of these early electron synchrotrons probably did not produce research results of the highest importance for the advance of scientific knowledge. The storage-ring work at Stanford—in particular in 1976 when it received an average of 25 citations per paper—and at Hamburg (8·7 citations per paper in 1978) had far greater impact than any of the research carried out on the original synchrotrons (Table XI). This conclusion is in line with the view of most of the physicists interviewed that electron high-energy physics only really came into its own with the advent of storage rings. Before this, proton accelerator physics had always been pre-eminent.

Many scientists argue that it is not the great bulk of papers that advance scientific knowledge most, but rather a small number of papers that exercise a disproportionately large impact on the growth of knowledge. These major advances are, by and large, represented by very frequently cited papers (Table XII). The top one per cent. of most-cited publications

AVAILABLE

TABLE XI

Average Citations per Publication for All Work Published in last Four Years Including Storage-Ring Work

Citations per publication in	CEA	DESY	NINA	SLAC
1966	5·8	—	—	—
1968	4·1	6·2	—	—
1970	3·0	4·6	—	8·1
1972	4·0	4·4	2·8	6·5
1974	7·6	3·1	2·3	4·7
1976	—	3·1 (7·2)[a]	1·9	8·5 (25·1)[a]
1978	—	5·7 (8·7)[a]	0·9	6·1 (12·2)[a]

[a] The figures in brackets are the citations per publication for the storage-ring work alone.

SOURCE: Manual analysis of data from *Science Citation Index.*

TABLE XII

Numbers of Most Frequently Cited Papers

Numbers of citations: in one year	Numbers of papers:[a] CEA	DESY	NINA	SLAC
12	19 (21)	24 (38)	1	33 (68)
15	10 (12)	15 (26)	0	26 (58)
20	4 (6)	4 (11)	0	12 (37)
30	0 (2)	1 (4)	0	5 (20)
50	0 (1)	0 (0)	0	2 (10)

[a] The figures in brackets are numbers of most frequently cited papers including those produced from work on storage rings.

SOURCE: Manual analysis of data from *Science Citation Index.*

arising from work on electron accelerators were cited 30 or more times in one year. Stanford users produced five such papers, Hamburg one, and neither Cambridge nor NINA succeeded in producing any. Indeed, NINA did not produce any papers cited 15 or more times in any one year, while Cambridge produced ten, Hamburg 15 and Stanford 26.

Evaluation by Peers

We asked about 80 high-energy physicists to specify the main contributions arising from work on their own accelerator—or on the one which they used for most of their research—and on a number of others, including all the major proton and electron high-energy physics accelerators. These same physicists were also requested to rank the accelerators in order of the merit of their contributions to science over the decade between 1969 and 1978.

There were slight differences between the physicists' rankings of work done on their own accelerator and the rankings by colleagues elsewhere of that work.[26] We regard the rankings made by the latter as the more valid indicator of the relative achievements of different accelerators. The high-energy physicists whom we interviewed, who, it should be stressed, were mainly electron high-energy physicists, believed the accelerators at Stanford contributed most to scientific knowledge over the decade in question; they placed Stanford significantly ahead of the large proton accelerator centres at CERN, and the two other American national laboratories, Fermi National Accelerator Laboratory (Fermilab) and Brookhaven (Table XIII).

We also asked the high-energy physicists to identify major changes in

TABLE XIII

Physicists' Evaluations of the Contributions to Scientific Knowledge made by Accelerators at 13 High-Energy Physics Centres between 1969 and 1978

(1=highest ranking; 13=lowest ranking)

	Evaluation by:				Average rankings (excluding self-rankings)
	Physicists at:			Other high-energy physicists	
	Hamburg (n=23)	Daresbury (n=16)	Stanford (n=16)	(n=21)	
Electron:					
Cambridge	8·4	8·1	8·1	8·6	8·3
Hamburg	3·9	5·0	4·0	3·7	4·2
Daresbury	7·3	7·0	8·5	8·1	8·0
Stanford	1·7	1·8	1·3	1·6	1·7
Yerevan	11·7	11·5	11·5	11·7	11·6
Proton:[a]					
Argonne	6·8	6·8	7·1	6·6	6·8
Brookhaven	3·4	3·4	3·6	4·3	3·7
CERN	3·0	3·1	2·4	2·5	2·7
Dubna	11·7	11·2	11·5	11·8	11·6
Fermilab	3·5	2·7	3·8	3·0	3·2
Moscow	12·0	11·5	11·5	12·0	11·8
Rutherford	8·2	9·2	8·3	7·6	8·3
Serpukhov	9·5	9·8	9·5	9·5	9·6

[a] This is a list of the major proton accelerators (accelerators whose energies were greater than about 7 GeV) that were operating during the ten-year period under consideration. Three of these are in the United States (Argonne, Brookhaven and Fermilab), three in the Soviet Union (Dubna, Moscow and Serpukhov), one in Britain (Rutherford) and one on the Swiss-French border (CERN).

---

[26] For example, NINA users gave themselves the ranking 7·0, while the average given by the other three groups of assessors was 8·0.

the relative positions of these centres over this period. In their view, at the beginning of the decade, the accelerators at Brookhaven and perhaps Geneva (CERN) were the world-leaders in high-energy physics. They thought that the leading position then passed to Stanford, largely because of the experimental work carried out on its storage ring. For the major part of the decade, the world-leaders in high-energy physics were considered by our reviewers to be the scientists working with the accelerators at Stanford, the European Organization for Nuclear Research (CERN) and the Fermi Laboratory at Weston, Illinois. They placed in the second rank a group of accelerators comprising those at Hamburg, Brookhaven (which was relegated to this second group when the Fermi Laboratory began to operate) and Argonne. Cambridge, Daresbury, Rutherford, and perhaps Serpukhov were placed by the high-energy physicists interviewed in a third division of accelerators, the scientific achievements of which were significantly smaller over the ten years than those by the users of centres in the second group. The relative positions of NINA, and of Nimrod, the proton accelerator at Rutherford Laboratory, were regarded as having remained fairly static during the period; neither was viewed as having ever risen above this third group during the time they were operating. Both NINA and Nimrod were, however, thought to have performed substantially better than the remaining Russian accelerators, which were placed in the fourth division.

Over the first half of the ten-year period, during which all four electron synchrotrons were operating, and the work of the greatest impact was being carried out, the users of Hamburg were thought to have contributed more to scientific progress than those of the Cambridge facility, and considerably more than those of NINA; the accelerator at Yerevan, according to the physicists we interviewed, contributed very little indeed. Taking the period from 1969 to 1978 as a whole, however, there is apparently little or no difference between the total contribution to scientific knowledge made by the Cambridge users (from whom the last publication appeared in 1974) and by research workers using the Daresbury accelerator; in both cases, the contributions seem to have been much greater than those from Yerevan, but significantly less than those of Hamburg.

The Significance and Limits of Assessments of Past Scientific Performance in the Making of Science Policy

Since Dr. Alvin Weinberg's initial discussion of the criteria for scientific choice, there has been little work carried out to answer such questions as "How well is a particular piece of research done?" and "How much impact on the advance of scientific knowledge does it have?" It is desirable to try to answer these questions in a systematic way.

In the case of our assessment of high-energy physics, we can conclude that the scientific achievements of the users of the Stanford and Hamburg accelerators appear to have been greater than those from the Cambridge and Daresbury synchrotrons, both of which seem to have made contributions of about equal value to high-energy physics during the decade in question. It should be borne in mind, however, that the Cambridge accelerator was closed in 1972 and was considerably less expensive than the machine at Daresbury. Moreover, work carried out on NINA yielded fewer experimental papers than its competitors and these were not on average much drawn upon by other scientists. Nor did NINA produce any very frequently cited papers. This finding is consistent with the directly expressed views of high-energy physicists about the absence of any major discoveries made at Daresbury. This is not to say, however, that no work of high quality was carried out on NINA. The work on tagged photon beams, and on combined polarised beams and targets, was regarded by high-energy physicists as being of high quality. But by the time this and other work at Daresbury was carried out, the interests of high-energy physicists had generally moved on to other areas, so the influence of this work was not as great as it might have been a few years earlier.

What, then, is the relevance, if any, of these results to science policy? When the decision was made to close NINA, no systematic study of the type we have undertaken was carried out. The decision was undoubtedly made on the basis of some form of assessment by eminent scientists. Such assessments could now be supplemented by studies of the sort we have undertaken.

If making major advances in scientific knowledge were the only criterion used in policy-making, then the decision to close down NINA was surely the correct one. Although this is only one factor to be considered in determining the allocation of scientific resources, it should be one of the most important considerations—especially in basic science where the primary objective is to contribute to the advance of human knowledge. Nevertheless, there are other issues to be considered; in particular, the Daresbury accelerator was what several scientists called a "second generation" electron synchrotron, and the aims of its users were inevitably restricted to improving on earlier work and carrying out more detailed experiments.

In addition, past performance may give only a limited picture of future work or promise. It might be argued that what NINA did, and what it could have continued to do had it been allowed to survive, was to produce definitive sets of data which, even if they did not have a particularly great influence at the time, would continue to be cited over a long period because of the accuracy of the measurements. We did not, however, find a great deal of support for this view among high-energy physicists, except among those who were users of NINA.

Some would claim that, even if it was never a world-leader in the field, the Daresbury accelerator nevertheless played a fundamental role in training British high-energy physicists, providing them with the opportunity to construct and use highly sophisticated equipment. In this respect, NINA might be seen as an "entry cost" to be paid so that British scientists, mainly in the north of the country, could join the world community of high-energy physicists. Many of the experimentalists from Glasgow, Lancaster, Liverpool, Manchester and Sheffield Universities currently using accelerators overseas, particularly at CERN and Hamburg, gained their first experience of high-energy physics on NINA. Whether or not they could have gained this necessary experience elsewhere, for example at the Rutherford Laboratory or at CERN, is a question we cannot answer. It should be noted, however, that in the discussions leading up to the original decision to build the Daresbury accelerator, there was little or no reference to the need for NINA as a training facility. This particular justification for the accelerator was only invoked at a later stage.

The past scientific performance of research groups or centres should never be considered in isolation. Nevertheless, in making policies which affect a basic science like high-energy physics, where the primary aim is to add to the sum of scientific knowledge rather than to create educational or technological benefits, past research performance should clearly be given great weight. In this paper, we have endeavoured to show that it is now possible to arrive at reasonably systematic assessments of the research performance of scientific groups that might be of use in the making of science policy.

• *ABSTRACT*

In a parallel paper, we have outlined a methodology for assessing the comparative scientific performance of large basic research facilities (and their associated user groups) working in the same specialty, and applied this method of 'converging partial indicators' to an evaluation of the contributions to science made by a number of radio telescopes. In this paper, we employ this methodology to evaluate the scientific performance of various optical telescopes — in particular, the 2.5-metre Isaac Newton Telescope, operated as a central facility by the Royal Greenwich Observatory in South-East England. For several years, this was Britain's only major optical telescope, as well as being the largest such instrument in Europe. We compare its performance over the last decade with that of three American telescopes of similar size.

This paper has three aims: first, to ascertain whether the method of converging partial indicators, originally applied to radio astronomy, provides a more general policy tool that can be extended to other specialties; second, to determine just how successful each optical telescope has been in producing new astronomical knowledge over the past decade; and, third, to discuss whether our results on the comparative scientific performance of the Isaac Newton Telescope may have any implications for British astronomy policy in general.

Assessing Basic Research: The Case of the Isaac Newton Telescope

John Irvine and
Ben R. Martin[1]

Astronomy has had a long and (in large part) distinguished history in Great Britain. Prominent in that history have been the Royal Greenwich Observatory (RGO) and the Royal Observatory, Edinburgh (in recent years operating the UK Schmidt and Infra-Red

---

Social Studies of Science (SAGE, London, Beverly Hills and New Delhi), Vol. 13 (1983), 49-86

Telescopes). Between them, these have been responsible for a major portion of British astronomical research, as well as providing central facilities for astronomers in a number of universities. However, in the first half of the twentieth century, although British pre-eminence in theoretical astronomy continued, the lead in optical astronomy was wrested from Britain by scientists in the United States, who exploited the large, modern telescopes constructed on the basis of generous donations from philanthropic industrialists. Immediately after World War II, in an attempt to restore British optical astronomy to its former glory, scientists put before the British Government proposals to build a magnificent new telescope, to be called the Isaac Newton Telescope (INT), second only in size to the giant 200-inch reflector soon to be completed on Mount Palomar in California.[2] The decision to build this instrument was widely hailed as 'the greatest single contribution to the development of observational astronomy ever made in this country',[3] and as such was seen as likely to provide a fitting national memorial to Britain's greatest scientist. Given that, until the mid-1970s, the INT remained the only large British optical telescope, and that it, along with the Royal Greenwich Observatory (which was entrusted with overall responsibility for its construction and operation as a central facility), absorbed a major fraction of Britain's optical astronomy resources in that period, it is pertinent to analyze precisely what returns have come from this large investment. What contributions to scientific knowledge has it been responsible for? Did its operation lead to any other major benefits — for example, instrumental innovation, or the training of young astronomers? In what follows, we shall attempt to provide answers to some of these questions.

In two other papers,[4] we have explained why there is an urgent and increasing need for systematic assessments of research performance in basic science. We pointed to three main reasons. First, expenditure on certain types of scientific research has escalated rapidly, particularly in the Big Sciences (high-energy physics, astronomy, and space research). Second, growth rates in the overall science budgets of Western nations have steadily declined, and, for the 1980s, as those nations wrestle with the problems of economic recession, the prospect is one of near-zero growth. Last, peer-adjudication, the method traditionally favoured by science policy-makers for establishing priorities, seems to be facing increasing difficulties as a tool for determining the distribution of scientific resources.[5] The problem is that, in Big Science specialties, there are typically only one or two major experimental facilities in each country. These are generally operated by central staff, who are responsible for most construction, technical development and service activities, but who also use the research facility for a significant proportion of the time, often in conjunction with external users. With research concentrated at such centres, a large section of the nation's scientists in the specialty concerned will have some political interest in all major capital projects proposed to the appropriate funding body for each central facility. This obviously has implications for the system of peer-adjudication normally used in determining whether to support new projects, since the effective operation of such procedures depends on the existence of a constituency of 'neutral peers' able to offer impartial advice on competing project proposals, and few, if any, such 'disinterested' peers may exist.[6] This problem may be all the more important in an economic climate where little growth can be expected in the overall science budget, since continued progress in science will depend more and more upon the existence of mechanisms capable of ensuring the rapid shift of resources from old, and no longer productive, areas of research, to promising new areas.[7] Consequently, it is essential to begin exploring other methods of obtaining the systematic information needed to complement traditional methods of determining scientific priorities. This was the overall objective of the project of which this study forms a part.[8]

What are the principal criteria that should be employed to establish scientific priorities? Weinberg was one of the first to address this problem, putting forward a number of internal and external criteria for scientific choice.[9] However, while he (and other writers who contributed to the subsequent debate) went some way towards providing a framework for assessing external benefits from scientific research (benefits such as trained manpower and technological spin-off, for example), there was considerably less progress in establishing criteria to be used in assessing internal benefits, and on how these criteria could be used in practice. Yet, for the reasons outlined earlier, it is becoming essential in today's changed economic climate that some consideration be given to the question of their systematic application.

The criteria suggested by Weinberg provide an appropriate starting point:[10] (I) Is the field ripe for exploitation?; (II) Are the scientists in the field really competent? As we have argued elsewhere,[11] the first of these is best separated into two parts; (Ia) What is the state of the field itself — is it ripe for exploitation?; (Ib) How well placed are those scientists requesting funding in relation to their likely competitors — do they possess some advantage (in time, or in instrumental capability, for example) enabling them to exploit the field ahead of such competitors? Of these three criteria, the first (Ia) is probably least amenable to systematic assessment. One can, of course, ask researchers in that scientific field for their opinions, but few are likely to admit that their field is not 'ripe' for exploitation (in the sense that *some* progress could always be made if they were allocated more money). Question Ia would indeed be better posed in the following terms: Is this field *more* ripe for exploitation than all other fields in which these scientific resources could be deployed? Then one can clearly appreciate the extreme difficulty of ever finding anyone in a position to answer it adequately.

For internal criteria Ib and II, however, more systematic data should be obtainable. For the latter, one of the most relevant pieces of information concerns the past research performance (the 'track-record' — that is, the magnitude of previous contributions to scientific knowledge) of the research group or instrument concerned; while for Ib it is clearly important to have reliable data on the track-record of the likely competitors. It must be emphasized that we are not claiming that track-record is the sole factor determining future research performance — merely that it is one of the most significant predictors. Undeniably, there are other factors, such as the field's 'ripeness': but, if these are so ill-defined that they cannot be evaluated in any systematic manner, then there is no alternative but to assume that a major new project stands a *greater* chance of being carried out successfully by researchers or institutions with a strong track-record over recent years, than by those with a poor record.

It is with the aim of assessing the scientific performance of central facilities and their user groups relative to that of their competitors within the same specialty, that we have put forward the method of converging partial indicators. The method can be summarized briefly. While there exist no perfect measures of contributions to scientific knowledge, there are several 'partial indicators' — that is, variables determined partly by the magnitude of the particular contributions, and partly by 'other factors'. For example, one of the possible indicators (the number of scientific papers produced by the users of each facility) reflects not only their output of scientific ideas, but also the publication practices of the researchers concerned, and, in particular, the emphasis placed on publications in determining access to funds and promotion prospects. Some of the partial indicators refer to the overall contribution to scientific progress made by the users of each facility (for example: the numbers of publications, citations, and highly cited papers; peer-evaluation rankings), while others take account of differing scales of research activity and reflect their 'productivity' (or output per unit input) (for example: the number of publications per researcher; the cost per publication; and so on). If these partial indicators are to yield reliable results on comparative research performance, then the influence of the 'other factors' must be minimized: the test that this condition has been met is *convergence* between the indicators. On the basis of other assessments of Big Science facilities in radio astronomy and high-energy physics,[12] we believe that significant results can be obtained provided the following precautions are observed: (1) the indicators are applied to user groups, rather than to individual scientists; (2) a range of indicators must be employed, each focussing on different aspects of a central facility's performance; (3) the indicators can only be meaningfully applied to *matched* facilities and user groups, comparing 'like' with 'like' as far as possible; (4) the indicators based on citations must be seen as reflecting the impact, rather than the quality or importance,[13] of the research work; (5) because of the imperfect or partial nature of the indicators, only in those cases where they yield convergent results can it be assumed that the influence of the 'other factors' has been kept small (that is, that the matching of the facilities and user groups has been largely successful), and that the indicators provide a reasonably reliable estimate of the contributions to scientific knowledge made using the various research facilities.[14]

Using this method, we were able to identify with a certain degree of confidence the world leaders, or 'first division' telescopes, within radio astronomy, whose research performance has been significantly better than that of a number of 'second division' instruments — which in turn, have been appreciably more successful than certain 'third division' telescopes.[15] We argued that, if this procedure was carried out for other basic science specialties, this would provide valuable information for policy-makers responsible for the distribution of resources *between* specialties. (If other factors were equal or indeterminate, they would normally prefer to support a first-division facility in one basic science specialty rather than a second-division facility in another.) What was not clear from our earlier work, however, was the extent to which the methodology constructed for the evaluation of radio telescopes would prove appropriate for the assessment of facilities in other specialties, particularly when those facilities were not amongst the world leaders for their field (as two of the four in the radio astronomy study proved to be). Hence, it is partly to assuage such doubts that we present below an assessment of the scientific performance of the large 98-inch (2.5 metre) Isaac Newton Telescope at the Royal Greenwich Observatory, compared with that of similar instruments operated in other countries. Although less expensive than high-energy physics, optical astronomy nevertheless absorbs considerable financial and scientific resources.[16] Within Britain, between 1967 and 1979, the main capital facility was the INT, although by the mid-1970s the country also had a half share in the 3.9 metre Anglo-Australian Telescope in the Southern Hemisphere, and is currently engaged in building a 4.2-metre telescope on La Palma in the Canary Islands. As astronomers themselves admit:

> Most of the telescopes that we use are very expensive and it is our responsibility to use them effectively. It seems reasonable from time to time to take a hard look at the productivity of those telescopes and to ask whether we are getting our money's worth in our use of them.[17]

Hence, in addition to testing the method of converging partial indicators on another specialty, the other main aim of this paper is to examine the scientific track-record of one such expensive facility and its user group, comparing it with that of various competitors. We should then be in a better position to reflect on the relative success of British post-war policy for optical astronomy. First, however, a short digression on the history and background to the Isaac Newton Telescope is necessary.

Post-war Astronomy Policy in Britain and the Isaac Newton Telescope

The INT has had a somewhat chequered history.[18] Originally conceived in 1946 as the project that would restore Britain to its former pre-eminent position in optical astronomy, it took 12 years from the time financial approval was obtained to agree upon a design,[19] and a total of 21 years to finish the project, although even then it was not available for use by university astronomers until the end of the following year. While we need not be concerned here with the reasons for the delays in building the telescope,[20] several of the more important decisions involved should be mentioned. Two were clearly taken in, or around, 1946. First, the astronomers concerned elected to build a completely new telescope rather than a copy of an existing instrument, even though the latter could have been built more cheaply and (perhaps more importantly, given the absence of a single major British telescope in the Northern Hemisphere) far more quickly. Second, it was decided that the INT should be operated as a central facility by the Royal Greenwich Observatory, rather than by a new 'Central University Observatory' — as Plaskett, Professor of Astronomy at Oxford University and the originator of the INT project, had proposed[21] — although the telescope would, of course, still be available to university astronomers. The third decision concerns the siting of the INT. Although the possibility of an overseas site was only seriously considered, in open and public discussion, from 1955 on, our evidence (see later) suggests that the decision to build the telescope in Britain was effectively taken much earlier, and that some astronomers actively supported such a move. Despite the evident unsuitability of the British climate for astronomical observations,[22] senior astronomers argued that the telescope should be sited at the RGO rather than overseas, on the grounds that only then could the close links between theoretical and observational astronomy — links which are essential to the health of both these fields, and of the universities — be maintained.

In the flood of enthusiasm for the exciting new project, and perhaps because of the ease with which government backing was obtained, astronomers made one crucial oversight. In the thirteen years between being given the financial go-ahead and the awarding of the construction contract in 1959, no systematic site-testing was carried out at the proposed location for the telescope. Techniques of site-testing were well established in the United States, and such a procedure was regarded there as absolutely essential in selecting the best available site for any major new telescope. It would have required neither a great deal of time, nor of money, to carry out similar tests in Britain to establish just how suitable the proposed location would prove.

The INT was officially inaugurated in 1967. Within a few years, the unsuitability of the site was fully revealed, and in 1975 the decision was made to move the telescope to La Palma. However, just as the costs of the original telescope rose from an initial estimate of £200,000 in July 1946 to a final cost of approximately £1 million, so the costs of moving and upgrading it have escalated from £1.76 million in 1975 to the best part of £7.5 million in 1980[23] — a sum far greater than the original cost of the telescope, even after allowing for inflation. The telescope has recently been moved, so now would seem to be an appropriate time to assess its scientific record. To what extent has the INT lived up to the claims made for it — claims used at the time to justify the sizeable capital investment? In 1957, Sir Richard Woolley, Astronomer Royal and Director of the RGO, predicted that the INT would 'do much to promote a vigorous resurgence in the practical as well as the theoretical problems of astronomy in this country'.[24] Has the INT been able to fulfill this promise? And, if not, what went wrong?

Our results are perhaps also of some relevance to another, wider set of policy issues relating to the British plans for a Northern Hemisphere Observatory (NHO) on La Palma[25] — and in particular the decision that the RGO should be responsible for the construction and operation of the new facilities. These stem from the fact that there has in recent years been some criticism concerning the overall role of the Royal Observatories in British astronomy.[26] In interviews with us, critics argued that the present institutional structure of British astronomy (with the Royal Observatories essentially being part of the same public body — the Science and Engineering Research Council — that controls the funding of astronomy) is not ideally suited for ensuring a flexible and successful national policy towards optical astronomy. They claimed that this has led in the past to an overconcentration of resources on the two Royal Observatories. Instead, they would prefer to see universities planning and operating their own central facilities (as they do at the American Kitt Peak National Observatory).[27]

In view of this latter criticism, it is clearly desirable to have systematic information on the research record of the INT. Success or failure in the research record of a large centralized facility will depend partly on the users (both internal and external), and partly on the effectiveness with which the organization responsible for its construction, instrumentation, and maintenance (in this case, the RGO) carried out its duties.

AVAILABLE

Some consideration was given to this latter issue in the early 1970s. The Science Research Council established a Northern Hemisphere Review Committee to plan the future of British optical astronomy in the Northern Hemisphere. Although the Committee was not unanimous in its conclusions, a majority[28] recommended that it would be in the best interests of British astronomy that the NHO be run, not by the RGO, but as a new 'third centre', independent of the two Royal Observatories, and giving university astronomers a much greater possibility of participating in all stages of planning, instrumenting and operating the centre. In making this recommendation, however, the members of the Committee did not have the systematic data (for example, on publications and citations) to justify, on scientific grounds, why a new centre should be given responsibility for the NHO. In part, this was because the INT had only been operating for a few years, and, although certain astronomers were already dissatisfied, this was rather early to make a formal study of how well the INT had operated under RGO control. Nevertheless, a majority of the Committee did feel that, while the RGO was capable of coping with the needs of more traditional astronomy, it was not so well placed to meet the demands associated with operating a large modern telescope. In addition, the Committee produced certain financial figures which suggested that the RGO, when compared with American observatories, was relatively expensive to operate. However, in the absence of data on the research performance of the INT, the ensuing debate over the Committee's report appeared, to outside observers at least, to be conducted more in political than scientific terms,[29] and eventually the Science Research Council decided not to implement the Committee's main recommendations.[30] In the final section of this paper, we discuss how the scientific assessment results presented here, had they been available at the time, might have affected the decision in 1974 on the organization of the NHO — a decision involving considerable resources, and affecting the whole future of British astronomy.

The INT and its Closest Competitors

In applying the method of converging partial indicators to assess the scientific output of a telescope like the INT, the first problem to be faced is that of finding similar instruments between which meaningful comparisons can be made. Whereas this was not too difficult in our earlier study of radio astronomy[31] — a specialty where there are just a few major facilities, each with a fairly stable research group concentrating its work almost solely on that facility — optical astronomy is characterized by a rather larger number of facilities, each drawing its users from a much wider scientific community. Indeed, some optical astronomers move between instruments frequently, seeking observing time at a number of observatories; the most eminent astronomers typically gain access to perhaps four or five major telescopes each year. Although a few major observatories are still operated by individual universities (or small groups of universities), most are run as national centres — and, indeed, most are open to use by foreign astronomers. Given the commitment of many researchers to more than one telescope, one can only hope to evaluate the output from a telescope (or set of telescopes), rather than that of a fixed research group. So which telescopes are most similar to the INT?

The main determinant of the research capability of optical telescopes is the aperture (or, more strictly, the light-gathering power). Hence it would be somewhat unrealistic to compare the scientific performance of the 2.5-metre INT with that of the 5-metre telescope on Mount Palomar, because of the significant difference in size. We have therefore chosen to compare the INT with the 3-metre Lick telescope on Mount Hamilton in California, the 2.1-metre telescope at the Kitt Peak National Observatory (KPNO) in Arizona, and the 1.5-metre telescope at the Cerro Tololo Inter-American Observatory (CTIO) in Chile (which is run by the same group of American universities that operates the Kitt Peak Observatory).[32] The latter two are, like the INT, national facilities. The Cerro Tololo telescope is rather smaller than the INT, but by including it in the study we should be able to gain some idea of the potential results that Britain might have obtained if it had elected in 1946 to build a telescope somewhat smaller than the INT overseas.[33] The Lick telescope is a university facility, but it also serves astronomers from a number of University of California campuses, and, therefore, in terms of the institutional and geographical dispersion of its user community, is not too different in nature from the INT. All four telescopes are of fairly conventional design; all were operating throughout the 10-year period from 1969 to 1978; and the three larger ones involved roughly similar capital costs (in the region of £1 million).

BEST COPY AVAILABLE

One immediate complication in comparisons between these telescopes arises from the fact that they form part of observatories which operate several telescopes, and carry out a variety of tasks apart from astrophysics. The problem is that for certain types of information (like annual operating costs), although it is simple to acquire data for the observatory as a whole, it is extremely difficult to disaggregate these for individual telescopes. In Table 1, we therefore list not only the main details of the four telescopes under comparison, but also those for the other telescopes at each of the observatories.

Input Measures

Before considering the outputs from these telescopes, and assessing their relative contributions to science, we must first examine the various inputs that have been used in creating those outputs. The aim here is not to derive precise figures on the costs of carrying out astronomical research; rather, our point is that, if there are significant differences in the level of inputs between the telescopes, then these clearly have to be taken into account in evaluating the outputs.

One obvious input is the overall cost of running the telescope each year. However, because each observatory employs very different accounting procedures (for example, as to what counts as 'telescope operations' and what as 'central facilities'), and because each operates several different research facilities, it is extremely difficult to obtain *comparable* figures for the four telescopes. We have therefore adopted the following procedure, which, although not totally unproblematic, can at least be applied *equally* to all four observatories.

Beginning with the figures for the total annual operating costs of each observatory, we have first subtracted from this an appropriate percentage for any work that does not involve (stellar) astrophysical research at the observatory. For Kitt Peak, this involved deducting thirty percent from the total observatory costs to allow for their solar, planetary and space work, and another twenty percent for various central facilities used by astronomers to process observations obtained on telescopes other than those at Kitt Peak. In the case of the RGO, thirty percent of the total observatory costs was deducted for the work on positional astronomy and providing

TABLE 1
Main Stellar Telescopes Operated by the Cerro Tololo, Kitt Peak, Lick, and Royal Greenwich Observatories, 1978

Centre	Telescope	Completion date	Approximate capital cost (for post-1945 telescopes)[a]
Cerro Tololo Inter-American Observatory (CTIO)	4-metre	1974	
	1.5-metre	1968	
	Yale 1-metre	1974[b]	
	0.9-metre	1967	
	Michigan Curtis-Schmidt 0.6/0.9-metre	1967[b]	
	Lowell 0.6-metre	1969[b]	
	0.4-metre	1965[b]	
	0.4-metre	1961[b]	
Kitt Peak National Observatory (KPNO)[c]	4-metre	1973	$11.10M
	2.1-metre	1963	$ 2.90M
	1.3-metre	1966	$ 0.68M
	0.9-metre	1966	$ 0.98M
	0.9-metre	1960	
	0.9-metre Coudé feed	1972	$ 0.57M
	0.4-metre	1963	$ 0.09M
	0.4-metre	1962	
Lick Observatory	3-metre	1959	$ 2.80M
	0.9-metre	1895	—
	0.9-metre refractor	1888	—
	0.6-metre	1964	$ 0.58M
	0.5-metre astrograph	1939	—
	0.3-metre refractor	1882	—
RGO	2.5-metre	1967	£ 0.96M
	0.9-metre	1934	—
	0.8-metre	1890s	—
	0.7-metre refractor	1890s	—
	0.7-metre refractor	1894	—
	0.3-metre refractor	1859	—

a. No attempt has been made to adjust those costs for inflation.
b. This is the date the telescope was moved to Cerro Tololo.
c. Kitt Peak also operates several solar telescope facilities.

BEST COPY AVAILABLE

TABLE 2
Approximate Annual Operating Costs[a] and Numbers of Astronomers Using each Telescope

		CTIO	KPNO	Lick	RGO
Total observatory operating costs	1970		£2.68M[d]	£0.37M	£0.70M[b]
	1974		£3.33M	£0.57M	£1.35M[b]
	1978	~ £2.1M	£4.77M	£0.90M	£2.00M[c]
Estimated costs of stellar astrophysics work at the observatory	1970		£1.34M[d]	£0.37M	£0.28M
	1974		£1.67M	£0.57M	£0.54M
	1978	~ £2.1M	£2.39M	£0.90M	£0.80M
Percentage of the stellar astrophysics costs apportioned to the particular telescope being assessed		9.8%	17.2%	82%	72%
Estimated annual cost of the telescope being assessed	1970		£0.23M[d]	£0.30M	£0.20M
	1974		£0.29M	£0.47M	£0.39M
	1978	~ £0.21M	£0.41M	£0.74M	£0.58M
Capital costs amortized over the life of the telescope	1970		£0.02M	£0.02M	£0.05M
	1974		£0.02M	£0.02M	£0.05M
	1978		£0.03M	£0.03M	£0.05M
Total annual costs	1970		£0.25M[d]	£0.32M	£0.25M
	1974		£0.31M	£0.49M	£0.44M
	1978		£0.44M	£0.77M	£0.63M
Appropriate number of users[e]	1970	40	60	45	45
	1974	65	65	45	100
	1978	80	90	47	70

a. These figures are approximate estimates only, obtained using the procedure described in the text. To convert dollars into pounds sterling, the exchange rates used are those quoted in the *National Institute Economic Review*, Vol. 86 (1978), Table 25.

b. These are the annual costs after subtracting the costs of the Radcliffe and Cape Observatories then operated by the RGO.

c. The overall RGO budget in this financial year was £3.45 million, but this includes a large manpower and capital investment in the new NHO facilities which is difficult to separate out from total costs. Using the figure for the costs in 1974 (the last year in which the NHO expenditure was kept separate from that of the RGO),

ST COPY AVAILABLE

TABLE 2 (continued)

and assuming a 50% increase in running costs in the intervening 4 years (the Retail Price Index in fact rose by 80% over this period, so this calculation probably underestimates the RGO costs), leads to a figure of £2.0 million for the estimated costs of the RGO in 1978.

d. The figure for the total costs of Kitt Peak in 1970 includes a large capital investment in the new 4-metre telescope. Since the capital expenditure in that year was almost certainly greater than the running costs of this telescope once it began operating, use of Abt's formula almost certainly results in an over-estimation of the costs of operating the 2.1-metre telescope in 1970.

e. These include PhD students.

---

certain national services (the Nautical Almanac and Time Services), and a further thirty percent for support services to telescopes overseas. The relevant figures are given in the second row of Table 2. It should be noted that these costs include expenditure on those non-telescope facilities (such as astronomical plate-measuring machines and computers) required for processing observations made with the telescopes at the observatory — expenditure which, we would argue, must be regarded as part of the total costs of operating a large modern telescope. The next stage is to apportion the total costs of astronomical work at each observatory between all the telescopes operated by it. In doing this, we have used the finding by Abt that, within KPNO, the annual operating costs of telescopes vary as the 2.1-power of the aperture.[34] Assuming that the same relationship holds approximately for telescopes within each of the three other observatories, we have calculated the percentage of the total astronomical costs that must be apportioned to the four telescopes under consideration here. As can be seen from Table 2, the figure varies from about ten percent for the 1.5-metre telescope at Cerro Tololo to 82 percent for the Lick 3-metre telescope. Using this percentage figure, we can then estimate the approximate total costs of supporting and carrying out research with the four telescopes.[35]

If Abt's 2.1-power cost/aperture relationship were also to hold *between* observatories, we should expect that the operating costs of the CTIO 1.5-metre telescope would be about half those of the KPNO 2.1-metre telescope, the INT to be 45 percent more expensive, and the Lick 3-metre telescope to be just over twice as expen-

BEST COPY AVAILABLE

sive, as the KPNO instrument. After allowing for the approximate nature of the estimates in Table 2, one can see that the costs are indeed roughly in these ratios.

In addition to annual operating costs, there are also the capital costs to be taken into consideration. Normally, big telescopes can be expected to have a long working life — perhaps fifty years or more.[36] If capital costs are written off over a period of fifty years, this adds only approximately £0.02 million[37] to the annual running costs of each of the American telescopes, making little difference to the figures quoted in Table 2. However, in the case of the INT, because this was only used by university astronomers for about 10 years (and by RGO staff for 12 years) before being refurbished and moved to La Palma, at a cost not greatly dissimilar from that of a new telescope (see note 23), its capital costs must be written off over a rather shorter period. We have used a figure of twenty years, on the grounds that the telescope was used in Britain for some ten years, and on the optimistic assumption that half of the telescope's components will be re-used on the new site. This adds £0.05 million to the annual costs of the INT. However, as the figures for the total costs on the bottom row of Table 2 indicate, this makes little difference to the *ratios* of costs for the four telescopes. Since (as was seen above) these ratios are roughly what one would expect on the basis of the relative sizes of the telescopes, one must conclude that in world terms the INT was *not* significantly less well supported than other similar telescopes. Any difference between its scientific performance and that of the American telescopes cannot therefore be attributed to a comparatively low level of funding.[38]

In our studies of radio astronomy and high-energy physics, a second input considered was the number of scientists engaged in research at each facility. However, as was mentioned earlier, it is impossible to identify a stable 'user community' for each of the optical telescopes. In the bottom row of Table 2, we give the approximate number of astronomers using the telescopes in particular years, but it must be appreciated that some of these astronomers were permitted only four or five nights of observing, while others had twenty or thirty. The former is more likely to be the case at the national observatories (that is, at CTIO, KPNO, and the RGO), while the latter applies to many of the University of California astronomers using the Lick telescope. For them, perhaps the majority of their research will be carried out on the Lick telescope, while the Kitt Peak 2.1-metre and Cerro Tololo 1.5-metre

BEST COPY AVAILABLE

telescopes might provide their users with a much smaller fraction of their total observing time in any one year. This would then explain why each of the three national-facility telescopes supports a significantly larger user community than that for the Lick telescope. Because of this widely differing pattern in the usage of the four telescopes, this particular input indicator will not be used in any of the comparisons that follow.

The Assessment of Contributions to Scientific Knowledge

In comparing the scientific output from research facilities, perhaps the most important data source is the scientific publications produced by researchers using those facilities. It is principally through publishing papers in scientific journals that researchers communicate new ideas and results to other scientists. Although there are undoubtedly other, less formal channels for disseminating research findings, it can generally be assumed that information passing through these channels eventually ends up in the scientific literature — otherwise, research publications would not form a key element in the reward structure under which scientists operate.[39] In our work, we have therefore taken numbers of publications as constituting an important, although partial, indicator of the magnitude of the contributions to our knowledge of astronomy made by different research groups. How, then, do the four telescopes with which we are concerned compare in their output of scientific papers?

In Table 3, we present data on the numbers of research papers published in refereed scientific journals each year that contain observational results obtained with the telescopes. It can be seen that the Lick and Kitt Peak telescopes have both produced an average of just over forty papers a year during the last decade, and the smaller Cerro Tololo telescope thirty-five. In contrast, the figure for the INT is seven. An even larger difference is shown in the figures on costs per paper; research publications from the Kitt Peak telescope have cost only just over one tenth those from the INT.

However, as we have been at great pains to point out in our previous work,[40] publication rates are but a partial indicator of the contributions made to scientific knowledge. It is possible that a

BEST COPY AVAILABLE

TABLE 3
Output Indicators for Optical Telescopes: Publications

		CTIO 1.5-metre	KPNO 2.1-metre	Lick 3-metre	INT 2.5-metre
No. of publications in[a]:					
	1969	13	22	33	1
	1970	15	29	29	0
	1971	27	33	35	6
	1972	35	43	30	5
	1973	39	38	40	14
	1974	51	54	37	6
	1975	46	51	44	9
	1976	45	56	51	10
	1977	44	43	62	10
	1978	39	65	61	10
Aggregate for 1969-78		354	434	422	71
Yearly average for 1969-78		35	43	42	7
Approximate cost per paper[b]:	1970	?	£7k	£10k	£68k
	1974	?	£6k	£11k	£53k
	1978	£6k	£7k	£13k	£63k

a. These figures are based on publication lists drawn up using documentation provided by the observatories and supplemented by various other data sources. Checking all the papers in randomly selected volumes of astronomical journals suggests that the publication lists obtained in this way are between 85% and 95% complete. Hence the figures in Table 3 should be accurate to within 10 or 15%, which is sufficient for our purposes here.

b. Because the number of papers published in any one year is relatively small, there are significant random fluctuations from year to year. In an attempt to 'smooth out' these fluctuations, we have taken an average of the numbers published over a three-year period, beginning with the year in question. Thus the 1970 figures are calculated by averaging the numbers of publications in 1970, 71 and 72, and dividing by the total costs of the telescope in 1970. Similarly the 1974 figures are based on the average number of publications between 1974 and 1976. (For the 1978 figures, however, we have used the numbers of papers published in 1978 only, since the 1979 and 1980 figures were not known at the time of completing the work reported here.) This procedure also allows for the delay (typically one year) between carrying out observations and publication of the results.

BEST COPY AVAILABLE

TABLE 4
Output Indicators for Optical Telescopes: Citations

		CTIO 1.5-metre	KPNO 2.1-metre	Lick 3-metre	INT 2.5-metre
Total number of citations to all observational papers published since 1969	1970	30	80	210	10
	1972	150	290	340	50
	1974	340	470	660	70
	1976	620	920	980	90
	1978	880	1200	1460	190
Number of citations to work published in the preceding 4 years only	1972	150	290	340	50
	1974	300	380	490	70
	1976	460	560	590	80
	1978	580	710	920	140
Average number of citations per paper for work published in preceding 4 years	1972	1.7	2.2	2.6	3.7[a]
	1974	1.9	2.3	3.4	2.1
	1976	2.6	2.8	3.4	2.0
	1978	3.3	3.3	4.2	3.6
OVERALL AVERAGE		2.5	2.7	3.5	2,8
Number of papers gaining n or more citations in one year	n = 12	21	31	41	4
	n = 15	16	20	23	2
	n = 20	4	5	13	1
	n = 30	0	1	4	0
Number of times highly cited papers received n or more citations in a year	n = 12	39	57	86	7
	n = 15	20	30	46	4
	n = 20	5	10	19	3
	n = 30	0	3	4	0

a. This figure is based on a relatively small number of papers (12, compared with about 100 for the three other telescopes), and the high value here is largely due to one very highly cited paper.

BEST COPY AVAILABLE

telescope with a relatively small publication output might have contributed significantly to the advance of knowledge if those publications have had a comparatively large impact on the scientific community. In this respect, we may regard the number of citations gained by all the papers from a telescope as providing a partial indicator of the total impact of that telescope's work. The figures for this have been obtained using the *Science Citation Index*, and are given in the first row of Table 4. They show that Lick papers have received between seven and ten times as many citations as those from the INT, the Kitt Peak papers between six and ten times as many, and the Cerro Tololo papers between three and seven times as many; these figures suggest that the total impact of all INT publications has been considerably less than that of even the smallest of the three American telescopes.

It might be argued that the INT did not really begin to contribute significantly to scientific knowledge until a year or so after the start of the ten-year period considered here. Although the telescope was officially inaugurated in 1967, it was not fully ready for use by university astronomers until some time later, and in fact produced only one paper in 1969 and 1970. One can, however, examine whether there have been appreciable changes over time in the impact of each telescope's work by looking at the citations to only the most recent work (in this case, work published during the preceding four years). The figures in the second row of Table 4 do reveal a significant increase in the impact of recent INT work — by a factor of nearly three between 1972 and 1978 — but there are roughly similar increases for all the three other telescopes,[41] so that the INT total for 1978 remains less than a quarter of the Cerro Tololo figure, less than one fifth the Kitt Peak figure, and less than one sixth the Lick figure.

What factors gave rise to these apparently very large differences in impact between the British and the American telescopes?[42] Undoubtedly the main factor was the substantially lower publication rate of the former. The figures on the average number of citations per paper earned each year by recent publications from the four telescopes suggest that, on average, INT publications have had a very similar impact to those from the Kitt Peak telescope, and probably slightly more impact than papers from the Cerro Tololo telescope. Lick papers, as might be expected given the greater size of the telescope, seem to have had the most impact of the four. If astronomers using the INT had published as many papers as their

BEST COPY AVAILABLE

colleagues using the American telescopes, then, provided that there was no 'dilution' in the impact of those papers, the overall impact of work performed on the INT might be expected to have been similar to that on the three other telescopes.

There is one further indicator based on citations that needs to be considered — namely, the number of highly-cited papers produced by each telescope. While the total number of citations, together with average citations per paper, may give a general indication of the impact of work from a particular telescope, it may not reveal whether that telescope has been responsible for any of the occasional 'discoveries' that play a major role in the advance of science. Scientific progress can come about through both incremental and revolutionary change, and it is conceivable that a facility having relatively low publication and citation rates might still be judged to have had a substantial impact on the advance of scientific knowledge if it could claim a few major discoveries. The distribution of such discoveries between the four telescopes assessed here, while it may be masked in the figures on total citations and citations per paper, should show up in data on the numbers of highly-cited papers.[43]

For these four telescopes, the top two percent most-cited papers gained twenty or more citations in one year. As can be seen from the penultimate row of Table 4, the Lick telescope produced thirteen such papers, the Kitt Peak telescope five, the Cerro Tololo telescope four and the INT only one. A similar picture emerges for papers cited twelve, fifteen or thirty times in a year; while, if allowance is made for some influential papers being longer-lived than others (bottom row of Table 4), the differences between the four telescopes are no less pronounced.

Finally, how do the results obtained using the various publication and citation indicators compare with those from peer-evaluation? To answer this, we asked about fifty astronomers (including a large fraction of the research staff of the Kitt Peak, Lick and Royal Greenwich Observatories, and some of their outside users) for their views on the relative successes and failures of the four telescopes considered here. We found, however, that many astronomers experienced difficulty in distinguishing work carried out on these telescopes from that performed on other telescopes at the same observatories. Consequently, the peer-evaluation results that follow refer to the overall output of all the main stellar telescopes at each observatory (those listed in Table 1),[44] although

BEST COPY AVAILABLE

some important non-quantitative data were obtained on the record of the individual telescopes with which we are primarily concerned.

Astronomers were invited to identify the main scientific contributions made by telescopes at their own observatory (or the one they use for the bulk of their research), and by those at eleven other observatories (including most of the world's major optical observatories) over the ten years between 1969 and 1978. They were then asked to place these in rank order (or, failing that, in 'first-class', 'second-class', 'third-class' and 'fourth-class' categories) according to the relative magnitude of their scientific contributions. Because the telescopes at twelve observatories were involved in this exercise, the rankings are on a scale between 1 (top) and 12 (bottom). The average rankings obtained from each group of astronomers are shown in Table 5.

As in our other surveys of radio astronomers and high-energy physicists, there was a high degree of consistency in the rankings obtained, both within each group of astronomers and between them. There were, however, small systematic differences between how astronomers ranked the output of the telescopes on which they worked (self-evaluation), and how their colleagues elsewhere ranked it (peer-evaluation).[45] Consequently, the average peer-evaluation rankings obtained after excluding self-rankings (these are shown in the final column of Table 5) should probably be regarded as the more reliable indicator. These show that observational work at Kitt Peak was rated 2.6 (second only to the Hale Observatories), at Lick Observatory 4.3, at Cerro Tololo 4.6, and at the Royal Greenwich Observatory 11.0. This suggests that optical astronomers judge the scientific contributions made using RGO facilities over the decade to have been very appreciably less than those made using the three American observatories.

Because these peer-evaluation data refer to all the main stellar telescopes at each observatory, rather than to individual telescopes, comparisons with results from other partial indicators must be approached with some caution. Publication and citation data all suggest that the Lick 3-metre telescope has contributed significantly more to scientific knowledge than the Kitt Peak 2.1-metre telescope, while the peer-evaluation results put the Kitt Peak work overall ahead of the Lick Observatory. The principal reason for this is likely to be that, since 1974, the main research instrument at Kitt Peak has been the 4-metre telescope, and it is this telescope, rather than the 2.1-metre instrument, that from then on has had the

BEST COPY AVAILABLE

TABLE 5
Rankings of Work Carried Out on the Telescopes at 12 Optical Astronomy Observatories Obtained by Peer-Evaluation

Observatory (size of main telescope in metres)	Kitt Peak scientific staff (n = 18)[a]	Lick scientific staff and users (n = 12)[a]	RGO scientific staff (n = 10)[a]	RGO users (n = 10)[a]	Average rankings (including self-rankings)[b] (n = 50)[a]	Average rankings excluding self-rankings[b]
Anglo-Australian Observatory (3.9)	4.6	5.1	2.3	2.8	3.7(±0.8)	3.7(±0.8)
Cerro Tololo (4.0)	4.8	4.5	4.9	4.0	4.6(±0.4)	4.6(±0.4)
European Southern Observatory (3.6)	6.6	7.0	7.6	7.6	7.2(±0.4)	7.2(±0.4)
Hale Observatories (5.1)	1.4	1.4	2.1	1.8	1.7(±0.3)	1.7(±0.3)
Kitt Peak (4.0)	1.9	2.8	2.5	2.5	2.4(±0.3)	2.6(±0.2)
Lick Observatory (3.0)	3.3	2.3	4.7	5.1	3.8(±1.1)	4.3(±0.8)
Mount Stromlo (1.9)	7.8	8.2	6.5	8.0	7.6(±0.8)	7.6(±0.8)
Haute Provence (1.9)	9.8	9.8	11.0	10.6	10.3(±0.5)	10.3(±0.5)
Royal Greenwich Observatory (2.5)	11.1	11.0	10.4	10.3	10.7(±0.4)	11.0(±0.1)
Russian 6-metre (6.0)	11.1	11.0	11.4	11.0	11.1(±0.1)	11.1(±0.1)
Radcliffe Observatory/ South African (1.9)	8.7	8.7	8.0	8.0	8.4(±0.4)	8.4(±0.4)
Texas McDonald Observatory (2.7)	6.7	6.2	6.6	6.3	6.5(±0.2)	6.5(±0.2)

[a] Sample size.

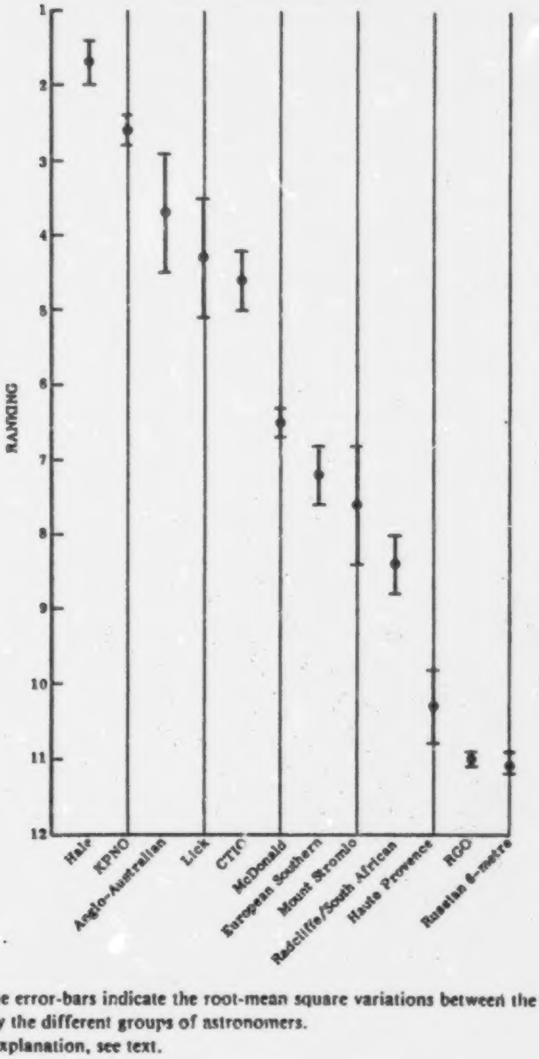
[b] The figures in brackets indicate the root-mean square variations between the rankings given by the different groups of astronomers, giving some approximate idea of the divergence in opinion of the different groups of scientists interviewed.

greatest influence in determining the overall scientific contribution of the observatory. In addition, there are several other smaller, but relatively productive telescopes at Kitt Peak, particularly the 1.2-metre instrument. If we had included the publications and citations stemming from observations on all these telescopes, the resulting Kitt Peak figures would have been considerably larger than those for Lick (even after adding in the work from the smaller Lick telescopes).[46] Similar considerations apply to Cerro Tololo, where the first papers from the 4-metre telescope were published in 1975. In this case, however, approximate calculations suggest that the figures on publications and citations would not have been very different from those for the Lick Observatory, in line with the very similar rankings that the two observatories were given in the peer-evaluation exercise. Finally, in the case of the optical facilities at the RGO, the INT has been the principal telescope throughout the decade. If papers from the smaller stellar telescopes at the RGO were included, this would probably add only about fifty percent to the publication and citation totals — leaving them a factor of six or more below the equivalent American totals. Overall, therefore, we see that, once consideration is given to the other telescopes at each of the four observatories, there is a reasonable degree of convergence between the results obtained with publication and citation indicators, and those based on peer-evaluation.

Before examining what conclusions can be drawn from this convergence, let us look in more detail at what the peer-evaluation results indicate. The results in the final column of Table 5 are shown pictorially in Figure 1, which illustrates more graphically the differences between the contributions from the telescopes at each observatory. What the figure and the table do not show, however, are changes in relative positions over time. In addition to ranking overall contributions to science over the decade, astronomers were asked to identify changes in relative positions over this period. It would seem from their responses that, at the start of the decade, the Hale and Lick telescopes were clearly perceived as the two world leaders in optical astronomy. However, once the new 4-metre telescopes at the Kitt Peak and Anglo-Australian Observatories came into operation in the mid-1970s, these two appear to have first caught up with, and then overtaken, Lick. Hence, many astronomers described the Hale, Kitt Peak, and Anglo-Australian facilities as making up the 'first-division' in optical astronomy. By 1978, Lick seems to have moved down to the 'second-division'

BEST COPY AVAILABLE

FIGURE 1
Average Rankings of Work Carried Out on the Telescopes at 12 Optical Astronomy Observatories for 1969-78 (excluding self-rankings)[a]



a. The error-bars indicate the root-mean square variations between the rankings given by the different groups of astronomers.

For explanation, see text.

BEST COPY AVAILABLE

because its 3-metre telescope found itself at an increasing disadvantage to the new 4-metre telescopes at observatories boasting significantly better observing conditions. Besides Lick, other members of the 'second-division' cited by astronomers were CTIO (perhaps now threatening to overtake Lick) and two university observatories not included in the peer-evaluation results — the University of Arizona Steward Observatory, and the University of Hawaii Observatory (whose position was also seen as improving rapidly).[47] The users of these were judged to have contributed significantly more to scientific knowledge than those of a number of 'third-division' facilities, including the Texas McDonald Observatory, the European Southern Observatory (where the new 3.6-metre telescope has so far failed to contribute as much as the other 4-metre telescopes constructed at about the same time), Mount Stromlo Observatory in Australia, and the South African Astronomical Observatory (whose main telescope was formerly at the Radcliffe Observatory). These, in turn, are seen as having contributed significantly more than several 'fourth-division' instruments, including the Haute Provence telescope in France, the Isaac Newton Telescope and the Russian 6-metre telescope.

In the particular case of work carried out on the INT, astronomers considered that its impact on the advance of knowledge over the ten years up to 1978 had been minimal, as compared with that of the major American telescopes, and many (including some at the RGO) had difficulty in pointing to any specific scientific contributions of importance. This cannot, in our view, be attributed to ignorance on the part of American astronomers of research work in Britain, since several were able to identify contributions made by the much smaller, 0.9-metre telescope at Cambridge University. Nor can it be blamed on chauvinism, since American astronomers rated the recent work of the Anglo-Australian Telescope quite highly.[48] It therefore seems difficult to escape the conclusion that astronomers see the INT as having made a relatively small contribution to science over the ten years between 1969 and 1978.[49]

Discussion and Conclusions

One aim of this paper was to determine whether the method of converging partial indicators could be applied beyond the specialty of

BEST COPY AVAILABLE

radio astronomy, where it was first used. Despite various difficulties in making direct comparisons, the partial indicators presented in this paper have yielded generally consistent results. The method appears to be capable of comparing the relative scientific contributions of optical telescopes, and of producing significant results. With the exception of the average number of citations per paper, the partial indicators seem to suggest that the Lick 3-metre telescope has been the most successful of the four, followed by the Kitt Peak 2.1-metre and the Cerro Tololo 1.5-metre, with the INT a considerable way behind. In terms of size, the scientific performance of the INT might be expected to be intermediate between that of the Lick and Kitt Peak telescopes. In fact, it falls short of this by a factor of six in terms of numbers of papers published over the decade, by a factor of seven in terms of the total number of citations gained in 1978, and by a factor of up to ten in terms of highly-cited papers. That the INT should appear to be so far behind the Kitt Peak 2.1-metre telescope (itself never more than a 'second-division' telescope) implies that its contribution to scientific knowledge has, in world terms, been rather small. If so, the INT has fallen well short of the claims made for it before it was built (and used to justify the major capital investment). This conclusion is shared by the great majority of the British astronomers whom we interviewed. Their disillusionment is mirrored in the rapid fall-off in the demand to use the INT; when it was completed, it was oversubscribed by a factor of four,[50] but by 1977 this had fallen to under 1.3, and our evidence suggests that, at times, the telescope was even undersubscribed.[51]

Before moving on to consider possible reasons for the relative lack of success of the INT, it is worth examining two other possible justifications that have been given in the past for its support: that it has proved a valuable training ground for young British astronomers; and that it has been an important test-bed for advanced astronomical instruments subsequently used on large overseas telescopes. To investigate the first, we asked the views of professors at four of the main university groups of INT users. All were sceptical of the value of the INT for educational purposes, suggesting that nothing is more disillusioning for young astronomers than to have their allocation of valuable observing time on a large telescope ruined by bad weather. They also reported considerable instrumental difficulties (one of the spectrographs, in particular, yielded unreliable results) which made it difficult for students to use the telescope.[52] Overall, it seems doubtful whether the use of the INT for training astronomers to operate a large telescope constitutes sufficient justification for its existence.[53]

The instrumental test-bed argument appears to be no more substantial. The most successful piece of optical astronomy instrumentation produced in Britain over the recent past — the Image Photon Counting System (IPCS) — was developed in a university astronomy department, rather than at the RGO. Although tested at the RGO, the small 30" telescope (rather than the INT) was used for its development, while the first astronomical observations with it were made at the Hale Observatories in California. According to the IPCS's designer, the INT, while used for some subsequent observations, played no significant role in this important instrumental development.[54] Yet, even if the educational and instrumental arguments for the INT had been as strong as some would claim,[55] we would still contend that investment in what was for several years Britain's principal optical astronomy facility had to be justifiable primarily in scientific terms. Our results suggest that, compared with other similar-sized telescopes, the INT proved relatively unsuccessful in this respect. What reasons can be found for this?

One possible explanation is that those astronomers using the INT were not as competent in undertaking significant research programmes as were their American counterparts. However, there is much cause to doubt this. American astronomers were in fact quite complimentary about British successes with the Anglo-Australian Telescope. Moreover, the many British astronomers (including some from the RGO) who have observed overseas have shown that, once given access to large telescopes on good observing sites, they are no less competent than other astronomers.

A second common argument is that the INT's relatively poor performance was due to inferior observing conditions in Britain. There can be no dispute that the RGO site proved totally unsuitable, confirming the worst fears of the few astronomers who warned against building the telescope in Britain *before* construction began. The site is at far too low an altitude — a mere 30 metres above sea-level, compared with several thousand metres for the best mountain sites. The humidity is often high (the RGO overlooks Pevensey Marshes), making frequent recoating of the telescope mirrors necessary. The street lights at a nearby town tend to illuminate the night sky. The INT was at one of the most

BEST COPY AVAILABLE

Northerly latitudes of all the world's large telescopes: in summer, when the weather is best for observing, the nights are rather short. Finally, and perhaps most importantly, the weather in Britain is far from good, with a high percentage of cloudy nights.[36]

The crucial question, then, is whether the site *alone* explains the factor (somewhere between six and ten) by which the scientific performance of the INT lagged behind that of equivalent telescopes overseas. According to the RGO there are, on average, approximately 1,200 hours of clear night sky at the site.[37] This is only a factor of two less than the 2,500 hours or so obtained in the South-West United States at Lick and Kitt Peak Observatories; this apparently leaves a factor of between three and five still to be explained. The number of hours of clear sky is by no means the only factor determining the astronomical observing conditions: but even if one allows for the additional effects of poor 'seeing' conditions, low atmospheric transparency, and the fact that a certain proportion of the quoted 1,200 hours of clear sky comes in the form of short periods (of little use for many types of observation), it is still uncertain whether the remaining factor can be completely accounted for — particularly since the Lick Observatory site suffers to a lesser extent from some of the same problems.

Many of the astronomers we interviewed were actually rather doubtful whether the INT's relatively poor performance could be explained in terms of the site alone, although they did accept that it was one major cause. They expressed concern with the way in which the INT and its associated instrumentation had been built and operated, and cited a number of what they perceived to be specific problems, about which university astronomers, in particular, clearly felt strongly. Those most frequently mentioned included relatively poor instrumentation (particularly the original coudé spectrograph, which was not ready until several years after the INT was completed, and which then proved unsatisfactory); the RGO's claimed failure to fulfill its role as a provider of central services (especially technical back-up) for university users of the INT, apparently because support staff at the RGO who were meant to be serving university astronomers were often too busy meeting the requirements of the RGO's own research staff;[38] a low level of instrumental innovation at the RGO compared with other major observatories overseas (the main reason for this being given as the isolation of the technical development work at the RGO from university astronomers); and a difference in approach towards the

BEST COPY AVAILABLE

demands of research — university astronomers contrasted what they termed the 'civil service' mentality of RGO staff with the approach found in their own departments (citing, for example, difficulties they encountered in attempting to call out RGO staff for assistance after hours). However, these criticisms are not accepted by most RGO staff. Although some admit that there may have been problems during the early years of the INT's life, when the RGO was only just beginning to adjust to its new role as a provider of national facilities,[59] they argue that significant advances have been made over the last few years. In their view, the perceptions of university astronomers have not reflected these changes, and the criticisms should therefore be seen as referring to a much earlier period in the INT's history.

There are thus two principal interpretations of the difference between the scientific performance of the INT and that of similar telescopes overseas. According to the first — the one favoured at the RGO — the poor performance of the INT is entirely due to the site. In contrast, according to a large section of the astronomical community outside the RGO, the site only explains part (albeit a major part) of the difference; the remainder is related to the way in which the RGO discharged its responsibility for the support and operation of the telescope. This latter interpretation appears to have been the one favoured by a majority of the Northern Hemisphere Review Committee at the time of their report.[60]

In the light of the INT's comparative lack of success, can any policy conclusions be drawn? First, subsequent events imply that the decision to build a 2.5-metre telescope in Britain was almost certainly wrong. For the same amount of money, a somewhat smaller telescope could have been built on a good site overseas. If this had been run by an organization similar to the group of universities operating KPNO and CTIO, then, in the opinion of many astronomers, it would have proved as successful as the Kitt Peak 2.1-metre telescope (which is about as far from the universities of some of its users as the La Palma site is from Britain) and the Cerro Tololo 1.5-metre instrument (which is far more remote). If the decision to build the INT in Britain was wrong, could this have been foreseen, and was there any alternative? The answer to the first of these questions is, 'Yes, almost certainly, if site-testing of the type already common in the United States had been carried out'. As for the second, many astronomers are now under the impression that the British Treasury and Admiralty, when they agreed to fund the

BEST COPY AVAILABLE

INT in 1946, would not have consented to an overseas site. However, we can find no evidence in the relevant public documents for the late 1940s to support this view.[61] Instead, all the documentary evidence indicates that the British site was chosen in 1946 by senior astronomers, who produced various arguments to justify this case. Only in the 1950s did the belief emerge among astronomers that the Treasury might object to an overseas site, *long after the original decision on the site appears to have effectively been made*. The threat that the Treasury might withdraw its backing for the project was then used to dissuade those younger astronomers who harboured doubts about the value of a British-based INT from expressing them in public.

The second set of policy implications concerns the decision to give the RGO responsibility for the new La Palma Observatory. If, as certain members of the astronomical community maintain, the RGO's handling of the INT, particularly with respect to their responsibility for providing modern instrumentation for the telescope, was indeed one of the reasons for the telescope's comparatively poor scientific performance, then some other justification apart from track-record with the INT must be sought for handing over responsibility for the new observatory to the RGO. However, because the report of the Northern Hemisphere Review Committee (NHRC) was not made available to the wider scientific community, there was no public debate as to whether such a justification actually existed.

Was there an alternative to making the RGO responsible for the La Palma Observatory? The NHRC certainly believed that there was. They argued forcibly for a 'third centre', funded by the SRC, similar to that operating the national astronomy facilities in the United States. But were there scientists of sufficient calibre to staff such a centre? Again the NHRC had no doubts. In addition to certain astronomers in the United Kingdom, some of whom have subsequently enjoyed great success using the Anglo-Australian telescope and other major instruments overseas, there were many British astronomers working in the United States, some with much distinction,[62] who at the time expressed a willingness to consider returning to Britain to take part in such a venture. The impressive nature of their track-record would certainly have argued well for the success of the La Palma Observatory. However, systematic, comparative data on the track-record of the INT were not available at the time. If information of the type described in this paper had been obtained at some stage between the completion of the NHRC

BEST COPY AVAILABLE

report and the eventual decision in 1974 to make the RGO responsible for the La Palma Observatory, then surely that report — and in particular the recommendation of a new independent centre to operate the observatory — would have received some public discussion, and its merits and drawbacks would have been openly assessed. The provision of information on the INT's performance up to 1974, even if in the last resort it had not led to a different decision, would at least have resulted in a *more informed* decision. Our results also appear to provide support for recent moves to ensure that the operation of the La Palma Observatory allows increasing scope, in its control and management, to university astronomers — in line with official SRC (and SERC) policy for the past fifteen years.[63]

• *NOTES*

The authors wish to acknowledge the support of the UK Social Science Research Council (grant No. HR 5175/1) in carrying out this research. They are also grateful for helpful comments on an earlier draft of this paper from various colleagues, astronomers and referees, and for a significant editorial contribution from David Edge.

1. No order of seniority implied (rotating first authorship).

2. The British Treasury approved plans for a 100-inch telescope, but the INT was in fact built with an aperture of 98 inches, slightly smaller than the 100-inch reflector on Mount Wilson, in California. By the time the INT was completed, another large (120-inch) reflector had been completed in California (at the Lick Observatory), so the INT was then only the fourth largest telescope in the world; however, it remained the largest in Western Europe until the late 1970s, when it was dismantled and moved to the Canary Islands. For a detailed history of the INT, see F.G. Smith and J. Dudley, 'The Isaac Newton Telescope', *Journal for the History of Astronomy*, Vol. 13 (1982), 1-18.

3. D.H. Sadler, 'The Isaac Newton Observatory', *Observatory*, Vol. 66 (1946), 380-83.

4. B.R. Martin and J. Irvine, 'Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, in press; Martin and Irvine, 'Internal Criteria for Scientific Choice: An Evaluation of the Research Performance of Electron High-Energy Physics Accelerators', *Minerva*, in press. We will refer to these papers, below, as (respectively) Paper 1 and Paper 2. Where page numbers are cited, they refer to the mimeo versions obtainable from the authors.

BEST COPY AVAILABLE

5. Cf. J. Irvine and B.R. Martin, 'What Direction for Basic Scientific Research?', submitted to *Nature*. For recent discussions of peer review, see I.I. Mitroff and D.E. Chubin, 'Peer Review at the NSF: A Dialectical Policy Analysis', *Social Studies of Science*, Vol. 9 (1979), 199-232; S. Cole, J.R. Cole and G.A. Simon, 'Chance and Consensus in Peer Review', *Science*, Vol. 214 (20 November 1981), 881-86; and D.R. Hensler, *Perceptions of the National Science Foundation Peer Review Process: A Report on a Survey of NSF Reviewers and Applicants* (Washington, DC: National Science Board and Committee on Science and Technology, US House of Representatives, December 1976).

6. The general implications of this for resource allocation in basic research are considered in Irvine and Martin, op. cit. note 5. For a brief discussion of 'research oligopoly' in British radio astronomy, see D.O. Edge and M.J. Mulkay, *Astronomy Transformed* (New York & London: Wiley-Interscience, 1976), 250-53. One possible consequence of the development of an 'oligopolistic' structure in Big Science is that funding decisions can at times appear to depend less on criteria relating to scientific merit than on political factors associated with the defence of institutional interests: it did not surprise us that several of the British astronomers we interviewed complained about what they saw as the control of research council committees by a secretariat who seemed to feel a greater responsibility towards SERC establishments than to university scientists. (The UK Science Research Council [SRC] was formed in 1965; it was renamed the Science and Engineering Research Council [SERC] in 1981.) Again, under certain circumstances, researchers associated with one Big Science centre may support applications for funds by other centres — even if it means failing to evaluate critically proposals of dubious scientific merit — because they recognize that it is in the common interest of all the centres in their specialty to maintain a united front vis-à-vis funding agencies. And, indeed, those we interviewed quoted various examples where they suspected that this had taken place. However, detailed examination of such claims is beyond the scope of this paper, and requires further careful research. What evidence there is suggests that such political pressures are present in most Big Science specialties, in most countries: the only exceptions would appear to be in cases where research councils seek the help of international experts in order to obtain impartial (or, at least, less partial) advice. For further discussion of these issues, see: D.S. Greenberg, *The Politics of Pure Science* (New York: New American Library, 1967), published in Britain as *The Politics of American Science* (Harmondsworth, Middx.: Penguin, 1969); M. Blissett, *Politics in Science* (Boston, Mass.: Little, Brown, 1972); M. Mulkay, 'The Mediating Role of the Scientific Elite', *Social Studies of Science*, Vol. 6 (1976), 445-70; S.S. Blume, *Toward a Political Sociology of Science* (New York and London: The Free Press, 1974); and C. le Pair, 'Decision-Making on Grant Applications in a Small Country', *Scientia Yugoslavica*, Vol. 6, Nos 1-4 (1980), 137-43.

As to decisions in British optical astronomy in particular, the situation is complicated since, from the late 1960s (following the formation, and developing influence, of the Science Research Council), official policy has been to develop the central, government-financed scientific facilities (such as the Royal Observatories) with more specific reference to the needs of university scientists. It was intended that the latter should play an increasing role in the planning and operation of these facilities. Whether, in astronomy, this policy has been successful will be discussed below. The sojourn of the INT at the RGO coincided with the gradual implementation of this policy, when the RGO was 'in transition'. This paper can be read as a contribution to a more pointed agenda for future research and discussion on these aspects of UK science policy.

7. Our enquiries suggest that, in at least one European country, there are already doubts as to whether peer-adjudication alone can guarantee the necessary flexibility.

8. The project aimed to assess the scientific performance of five major research facilities supported by the UK Science Research Council — the Royal Greenwich Observatory, the radio astronomy observatories at Cambridge and Jodrell Bank (see Paper 1, op. cit. note 4), and the Rutherford and Daresbury high-energy physics laboratories (see Paper 2, op. cit. note 4).

9. A.M. Weinberg, 'Criteria for Scientific Choice', *Minerva*, Vol. 1 (1963), 159-71. This paper led to a debate, the main contributions to which have been collected together in E. Shils (ed.), *Criteria for Scientific Development: Public Policy and National Goals* (Cambridge, Mass.: The MIT Press, 1968).

10. Weinberg, op. cit. note 9, 163.

11. Paper 2, op. cit. note 4, 12.

12. See Papers 1 & 2, op. cit. note 4, and note 8.

13. For a discussion of these terms, see Paper 1, op. cit. note 4, 15-17.

14. The details of how the main problems with the various partial indicators are overcome in the method of converging partial indicators can be found in Paper 1, op. cit. note 4, Figure 2, 25.

15. See J. Irvine and B.R. Martin, 'A Methodology for Assessing the Scientific Performance of Research Groups', *Scientia Yugoslavica*, Vol. 6, Nos 1-4 (1980), 83-95. Further comment is perhaps necessary here on our use of the term 'research facility'. In assessing the research performance of a facility like a major telescope, one is analyzing the relative magnitude of *all* the contributions to scientific knowledge coming from that telescope, regardless of whether it was used by the staff of the centre responsible for its operation, or by visiting users. The research performance of a facility therefore obviously depends on its use by both internal and external users, as well as on its quality, size and vintage, and the instrumentation and back-up services provided by the central staff.

16. The RGO and the Royal Observatory at Edinburgh have together accounted for a steadily increasing percentage of the total budget of the UK Science Research Council. The figure has risen from about two percent in the 1970-71 financial year to about four percent in 1978-79. To this must be added grants to university astronomers and subscriptions to various telescope facilities overseas.

17. H.A. Abt, 'The Cost-Effectiveness of Telescopes of Various Apertures', in A. Hewitt (ed.), *Optical and Infrared Telescopes for the 1990s* (Tucson, Ariz.: Kitt Peak National Observatory, 1980), 609-15, quote at 609.

18. Much of the material in this section on the early history of the INT has been assembled by one of us for a forthcoming doctoral thesis: B.R. Martin, *The Evolution of British Optical Astronomy, 1945-75: A Case Study in Science Policy* (PhD thesis, Department of Liberal Studies in Science, Manchester University, in preparation). This thesis draws heavily from Public Record Office documents, and other publicly available material, as well as interviews with certain senior astronomers. See also Smith and Dudley, op. cit. note 2.

19. In 1958, the British Government temporarily withdrew its backing for the project, presumably exasperated by the long delays in settling the design, and by the soaring costs (by then some three times the original estimate). This delay, however, accounts for only one year out of the twenty-one it took to complete the telescope.

20. For details, see Smith and Dudley, op. cit. note 2.

21. H.H. Plaskett, 'Presidential Address', *Monthly Notices of the Royal Astronomical Society*, Vol. 106 (1946), 80-94, esp. 93.

22. Of course, British astronomers always *knew* that, in terms of 'seeing' conditions, there were much better sites abroad, but it appears that, in the mid-1940s, many simply did not press to establish whether these were real alternatives, 'open options': see note 33 below, and the 'Discussion and Conclusions' section of this paper. As R.O. Redman (one of the strongest advocates of overseas sites) put it in 1946, 'Astronomers acquiesce, because the choice is between a new instrument in a bad situation or no instrument at all': *Observatory*, Vol. 66 (1945), 153-59, quote at 157.

23. This figure of £7.5 million for setting up the INT on La Palma now includes the cost of certain central facilities that will be used with the new 4.2-metre telescope, as well as with the INT. Even so, some of the astronomers we interviewed felt that this figure was rather high, and suggested that a new, purpose-built telescope could have been constructed for the same price. They argued that, since the new 4.2-metre telescope will cost some £9 million, and since the capital costs of telescopes vary approximately as the 2.4 power of the aperture (see H.A. Abt, 'The Cost-Effectiveness of Telescopes of Various Apertures' [Tucson, Ariz.: Kitt Peak National Observatory, mimeo, 1980]), such a new telescope could have an aperture of well over 3 metres — that is, appreciably greater than the INT's 2.5-metres. In contrast, other astronomers felt that the costs of the INT's move were justified: they seemed to welcome this opportunity (as they saw it) of, on the one hand, finally removing the embarrassment of a major research facility that was inevitably under-exploited because of the British climate and, on the other, leaving RGO staff free to concentrate on the new facilities on La Palma.

24. R. v.d. R.Woolley, as reported in *Nature*, Vol. 180 (31 August 1957), 420.

25. This is now known as the La Palma Observatory, or the Roque de los Muchachos Observatory, and will consist of the INT, the 4.2-metre telescope referred to in note 23, a 1-metre telescope, the British millimetre dish, and a number of smaller telescopes.

26. For example, see G. Burbidge, Correspondence, *Nature*, Vol. 239 (8 September 1972), 117-18; P.A. Strittmatter, Correspondence, *Nature*, Vol. 239 (20 October 1972), 475.

27. Others claim that this already effectively happens, through the RGO, following the change in policy starting in the late 1960s (see note 6).

28. Of the nine Committee members and consultants, just two (the Directors of the two Royal Observatories) dissented from this recommendation, which was itself an aspect of a general shift in central policy (see the end of note 6). It should be noted that this information, as well as later discussion on the role and recommendations of the Committee, has been drawn directly from the *Report of the Northern Hemisphere Review Committee*. Though this was never published by the SRC, we were given access to the Report by one of the members of the Committee. A detailed historical analysis of decisions in UK astronomy policy since the mid-1960s will obviously have to consider the precise composition of this Committee, and of other related bodies.

29. See, in particular, 'Scandalous Muddle in British Astronomy' (Editorial), *Nature*, Vol. 239 (15 September 1972), 121-22. It may have been partly in reply to the implication contained in the Northern Hemisphere Review Committee's report

BEST COPY AVAILABLE

(which, although unpublished, was the subject of well-informed rumour) that RGO astronomers had been less than successful as research scientists, that the Science Research Council claimed that 'the scientists working in the SRC Establishments' programmes are making *a significant contribution to the advancement of knowledge* in different fields': UK Science Research Council, *Annual Report for the Year 1971/72* (London: HMSO, 1972), 52 (our emphasis). Neither side, however, produced reliable data to back up their assertions.

30. From our interviews, it is clear that there was considerable resentment within the scientific community about the way the Committee's report was treated. The report was never made public, so denying the main body of the astronomy community the possibility of open discussion of its analysis and contents. Indeed, the report was kept so secret that one of the members of the Committee that had drawn it up informed us that, when he requested a copy, he was told that this was not possible because its contents were 'confidential'.

31. See Paper 1, op. cit. note 4.

32. There are several other telescopes of similar size, like the Steward Observatory 2.3-metre, and the Texas McDonald 2.7-metre and 2.1-metre telescopes, but these are all at essentially single-university observatories, while the four telescopes included in our sample are multi-university facilities.

33. Some astronomers (whatever their views — if any — may have been in the mid-1940s) now argue that there was little possibility of the British Government agreeing, in 1946, to build a large telescope overseas. However, Public Record Office documents of the period reveal that senior astronomers actually *wanted* the telescope to be in Britain (for the reasons we have already mentioned). In none of the documents for this period did they ascertain whether or not the Treasury would agree to an overseas alternative. For further details, see Martin, op. cit. note 18. (Telescopes sited abroad obviously cost more to operate, so a slightly smaller telescope would have had to be built in order to keep within the INT's budget.)

34. See Abt, op. cit. note 23.

35. It must be emphasized that the figures in Table 2 are intended to indicate total costs, rather than marginal or direct costs. Abt (op. cit. note 23, 15) quotes a figure of $173,000 (£74,000) for the direct costs of the Kitt Peak 2.1-metre telescope in 1974, at the same time pointing out that this represents only a small fraction of the total costs — it corresponds to one quarter of our estimate of the total costs. Similarly, material provided by the RGO suggests that the marginal costs of the INT in 1972 were about £100,000, again approximately a third or a quarter of the total costs, as estimated in Table 2.

36. In his calculations, Abt (op. cit. note 23) uses a figure of 75 years, but this would make little difference to the total costs in the final row of Table 2.

37. If one allows for the effects of inflation, the figure would be rather higher.

38. This is not to say that users of the INT were satisfied with the funds available for operating the telescope: many clearly were not. Rather, the conclusion is that, in this respect, their position was not significantly worse than that of their American counterparts. This is consistent with the view expressed in the Northern Hemisphere Review Committee's report that the RGO 'enjoys resources large in comparison with most observatories anywhere in the world'.

39. In the particular case of optical astronomy, researchers often use published papers to account for the periods of observing time they are allocated on major telescopes. This is probably why Abt (op. cit. note 23) finds that the number of

BEST COPY AVAILABLE

publications in any one year from each Kitt Peak telescope corresponds approximately with the number of users of that telescope in the previous year.

40. See, especially, Paper 1, op. cit. note 4.

41. These increases are not without significance for policy-makers. In our study of high-energy physics electron accelerators (Paper 2, op. cit. note 4), we found a rapid decline in the numbers of citations to recent work after the first few years of an accelerator's life, suggesting swift obsolescence of such machines as frontier-research tools. In optical astronomy, the reverse appears to be true: optical telescopes seem to become more productive with time (at least for the first ten years or more), presumably because better instruments are being fitted to them. This conclusion is in line with the long working life (fifty years or more) of telescopes, referred to above; in contrast, an accelerator has a useful working life of perhaps ten or fifteen years, after which it normally ceases to contribute significantly to the advance of high-energy physics, and is superseded by more powerful accelerators.

42. Those readers tempted to reply 'The British weather!' are referred to our later 'Discussion and Conclusions' section.

43. Peer-evaluation, in contrast, should provide some sort of weighted average for both the great mass of relatively minor contributons made by a telescope, and the small number of major discoveries: see Irvine and Martin, op. cit. note 15, 86.

44. In doing so, of course, we are temporarily shifting the focus of our quantitative evaluation from the single telescope at each centre included in our study, to *all* the stellar telescopes at each.

45. Kitt Peak astronomers ranked themselves 1.9, compared with 2.6 given by other astronomers; Lick, 2.3 compared with 4.3; and RGO users, 10.4 compared with 11.0.

46. In 1977, a total of approximately 140 papers were published using observations obtained on Kitt Peak telescopes — about twice the number produced from the various Lick telescopes. Using Abt's finding (op. cit. note 23) that the number of citations earned by Kitt Peak telescopes of different apertures varies as the 1.5-power of the aperture, one can estimate approximately the total number of citations that work published since 1969 from *all* the Kitt Peak telescopes would have obtained. In 1978, the total number of citations would probably have been over 3000, compared with well under 2000 for work from all the Lick telescopes.

47. These two centres were not included in the original list that we asked astronomers to rank. However, several of those interviewed mentioned these two observatories, generally placing them behind Cerro Tololo, and ahead of the University of Texas McDonald Observatory.

48. In any case, American astronomers' ranking of work at the RGO (11.0) is little different from that of British astronomers (10.3).

49. Some astronomers now find this conclusion so obvious that they wonder why we have felt it necessary to devote so many pages to reaching it. As one referee put it: 'Assessing the cost-effectiveness of the INT is like being asked to prove that it is colder at the poles than the equator'. However, having clearly summarized astronomers' opinions on this matter, he went on: 'but opinion does not persuade the powers that be, and consequently an exercise such as that represented by this paper is necessary for that purpose'.

50. That is to say, astronomers applied for four times as many nights of observing time as were available.

BEST COPY AVAILABLE

51. Evidence for this came from an interview with a former member of the SRC Panel for the Allocation of Telescope Time.

52. The head of one major university astronomy department recalled: 'We tried to use the INT once for a student thesis, but he nearly failed because it was impossible to get any reliable data. We never tried again.'

53. One senior astronomer commented: 'For many years Sir Richard Woolley, then Director RGO, would not allow astronomers time on the INT unless they had previous big telescope experience — a chicken and egg situation'.

54. While the INT has been used to develop various instruments and techniques, it is by no means clear whether it was *essential* for their development. Several astronomers stated that they used the INT for instrumental development solely 'because it was there', and because it was easy to obtain time on it; without the INT, the development would, they believed, have gone ahead on other telescopes, as it did to a large extent with the IPCS.

55. The same remarks apply to arguments based on technological spin-off from the INT. These have not been examined here, but our view is that the level of spin-off from the INT has been small compared with the capital investment involved, and also small compared with the spin-off noted at other Big Science centres in Britain: see J. Irvine and B.R. Martin, 'The Economic Effects of Big Science: The Case of Radio Astronomy', *Proceedings of the International Colloquium on Economic Effects of Space and Other Advanced Technologies*, Strasbourg, 23-30 April 1980 (Paris: European Space Agency, ref. ESA SP-151, 1980), 103-16.

56. Just how unsuitable the site is could have been established long before construction of the telescope began in 1959. Perhaps most senior British astronomers thought such a systematic site test otiose — a matter of 'proving it is colder at the poles than the equator'. The site for the La Palma Observatory was, of course, selected after systematic testing.

57. This is an average of the figures quoted in the RGO *Annual Reports* for the years between 1973 and 1978.

58. We cannot stress too strongly that we are *not* implying any judgment as to whether our respondents' perceptions of these 'problems' are fair and accurate — that, as we have already said (see note 6), requires further careful, detailed and specific research, and is beyond the scope of this paper. We are merely stating that these claims and assessments represent the *feelings* of many university astronomers.

59. See the end of note 6. As one senior astronomer, with experience in both universities and Royal Observatories, wrote to us late in 1981: 'The move from concentration on in-house research to the provision and operation of national facilities in the late 1960s and early 1970s involved a profound change in the Royal Observatories, and I think it is true to say that to a greater or lesser extent they resisted or went along grudgingly with the new regime. I know for a fact that there are still members of the Observatories who believe that the current role of the Observatories is wrong and that the old system was much preferable. However, the new role of the Observatories is now firmly established and the staffs of both Observatories are responding with enthusiasm to the new challenges.'

60. See note 28: as we state there, further research into the composition of this Committee, and into its findings, is required.

61. We have examined all the relevant documents available at the Public Record Office, the Ministry of Defence, and the Treasury. We have seen nothing to indicate that the expenditure of the same sum of money on a similar telescope overseas was

not an open option in 1946. For a contrasting interpretation of these events, see Smith and Dudley, op. cit. note 2.

62. Currently one such expatriate astronomer is Director of the Kitt Peak National Observatory, another is Director of the Steward Observatory (one of the most successful university facilities — see note 47), a third has been one of the most distinguished users of the Lick 3-metre telescope, and a fourth has figured prominently in the achievements of the Hale Observatories (by our peer-evaluation, the world's premier optical astronomy centre).

63. See the end of note 6. The so-called 'Tiger Team' set up to examine the RGO plans for the new 4.2-metre telescope contains a strong representation of university astronomers: this is another indication of the progressive development of this policy. Another is the fact that both the RGO and the Royal Observatory, Edinburgh, now have active young Directors who were previously university astronomers with international reputations — rather than civil service scientists already 'within the system'.

John Irvine and *Ben R. Martin* are Fellows of the Science Policy Research Unit, University of Sussex, where they work on a range of issues concerned with policies for science and technology. They are currently engaged on a project assessing The European Centre for Nuclear Research (CERN) and have published a number of articles in the area of research evaluation. John Irvine is the co-editor of *Demystifying Social Statistics* (Pluto Press, 1979) and *The Poverty of Progress: Changing Ways of Life in Industrial Societies* (Pergamon Press, 1982). *Authors' address:* Science Policy Research Unit, The University of Sussex, Mantell Building, Falmer, Brighton, Sussex, BN1 9RF, UK.

BEST COPY AVAILABLE

Assessing basic research

Some partial indicators of scientific progress in radio astronomy

Ben R. MARTIN and John IRVINE *

Science Policy Research Unit, University of Sussex, Brighton BN1 9RF, UK

Accepted for publication September 1980

As the costs of certain types of scientific research have escalated, and as growth rates in overall national science budgets have declined, so the need for an explicit science policy has grown more urgent. In order to establish priorities between research groups competing for scarce funds, one of the most important pieces of information needed by science policy-makers is an assessment of those groups' recent scientific performance. This paper suggests a method for evaluating that performance.

After reviewing the literature on scientific assessment, we argue that, while there are no simple measures of the contributions to scientific knowledge made by scientists, there are a number of 'partial indicators' - that is, variables determined partly by the magnitude of the particular contributions, and partly by 'other factors'. If the partial indicators are to yield reliable results, then the influence of these 'other factors' must be minimised. This is the aim of the method of 'converging partial indicators' proposed in this paper. We argue that the method overcomes many of the problems encountered in previous work on scientific assessment by incorporating the following elements: (1) the indicators are applied to research groups rather than individual scientists; (2) the indicators based on citations are seen as reflecting the impact, rather than the quality or importance, of the research work; (3) a range of indicators are employed, each of which focusses on different aspects of a group's performance; (4) the indicators are applied to matched groups, comparing 'like' with 'like' as far as possible; (5) because of the imperfect or partial nature of the indicators, only in those cases where they yield convergent results can it be assumed that the influence of the 'other factors' has been kept relatively small (i.e. the matching of the groups has been largely successful), and that the indicators therefore provide a reasonably reliable estimate of the contribution to scientific progress made by different research groups.

In an empirical study of four radio astronomy observatories, the method of converging partial indicators is tested, and several of the indicators (publications per researcher, citations per paper, numbers of highly cited papers, and peer evaluation) are found to give fairly consistent results. The results are of relevance to two questions: (a) can basic research be assessed? (b) more specifically, can significant differences in the research performance of radio astronomy centres be identified? We would maintain that the evidence presented in this paper is sufficient to justify a positive answer to both these questions, and hence to show that the method of converging partial indicators can yield information useful to science policy-makers.

1. Introduction – The need for an assessment of basic research

Few would now dispute the necessity for some assessment of the output from basic research. As the scale of basic scientific research has grown, so the need for an explicit and systematic science policy has increased - a policy that can only be sensibly constructed on the basis of an evaluation of the output from that research. When basic research was little more than a hobby for gentlefolk, and even as recently as 50 years ago when the annual equipment budget for an entire university department like the Cavendish Laboratory at Cambridge University amounted to only a few hundred pounds, it was entirely acceptable to leave questions concerning the organization and funding of science to scientists. There was no need for a policy establishing priorities across the sciences, nor for an evaluation of the performance of scien-

---

* No order of seniority implied (rotating first authorship). The authors are currently Fellows of the Science Policy Research Unit where they work on a range of issues connected with policies for basic research. They wish to acknowledge the support of the Social Science Research Council in carrying out this research. They also want to thank their colleagues at the Science Policy Research Unit, particularly Geoff Oldham, Keith Pavitt, and Roy Turner, for their help and comments. The conclusions are, however, the responsibility of the authors alone, and are not necessarily shared by their colleagues. Norman Dombey wishes to dissociate himself completely from the results of this publication.

Research Policy 12 (1983) 61–90
North-Holland Publishing Company

tists; any regulation of scientific affairs could be left to the scientists themselves. Now, however, the resources devoted to basic research are so substantial that its practitioners can no longer hope to escape the scrutiny of those seeking to know whether public funds are being spent in areas that yield a good rate of return (whatever form those returns may take). In Britain, the budget of the Science Research Council (SRC) for the financial year 1978–1979 amounted to well over £150 million. Although some of this was allocated to engineering research, and some to certain, more applied areas of biology, chemistry, and physics, the great bulk went to basic research – that is, research carried out with the primary purpose of increasing scientific knowledge rather than creating technological, social, or economic benefits.

The need for a policy for basic research, and hence for analytical tools to assess such research, has been greatly sharpened by the problems of economic recession in the 1970s. Industrialised nations have been forced to look more carefully at all areas of state expenditure. Whereas in the 1950s and 1960s economic growth rates generally permitted public expenditure in one area such as scientific research to grow, and even to grow quite rapidly, without affecting other areas, now an increase in one area of public expenditure tends only to be possible at the expense of cut-backs in others. Those scientists who work in high-cost specialties such as high-energy physics, space research, and astronomy – where new facilities can cost tens or even hundreds of millions of pounds – and who argue for continued real increases in their level of funding, must now be willing to accept wider public examination of their research work in order that one may ascertain what benefits are associated with it, and how these benefits compare with those from other areas competing for resources. It is no longer sufficient to assert that a particular project promises certain benefits or "good science"; instead, it must be demonstrated that the project is likely to yield *greater* benefits than any of the competing alternatives. Such a judgement can only be arrived at on the basis of an assessment of the benefits associated with all the various areas of basic research, and the establishment of a set of relative priorities between them in one comprehensive science policy.

For policy-makers, there are four main sets of decisions that need to be made in the allocation of resources to basic research.

(1) How much overall should be spent on basic research compared with other areas of public expenditure?
(2) How should this overall research budget be distributed between the different disciplines each with competing claims for funding?
(3) How much should be allocated to the different types of scientific institutions? In Britain, for example, what proportions should be spent on international facilities, on Research Council establishments, and on university research?
(4) How much should be allocated to each research centre, group, or individual, working within a discipline?

Some consideration was previously given to these questions in the debate on criteria for scientific choice during the 1960s (cf. Shils [49]). In particular, Weinberg [53, p. 163] proposed certain criteria on which decisions as to the distribution of resources to scientific research should be based, distinguishing between "internal" and "external" criteria:

> Internal criteria are generated within the scientific field itself and answer the question: How well is the science done? External criteria are generated outside the scientific field and answer the question: Why pursue this particular science?

In the debate stimulated by Weinberg's proposals, participants tended to concentrate more on the policy issues associated with applying *external* criteria to decisions about scientific funding, and hence to focus on the first two of the four questions identified above. Much attention was given to such factors as the contributions of science in general, or of a particular scientific discipline, to education, to technology, and to economic growth; and governments have in recent years begun to take account of some of these external factors in arriving at decisions on funding. However, internal criteria cannot be ignored, particularly in the case of more basic scientific research. It is with these that we are primarily concerned here.

In this paper, we present a framework for assessing the relative contributions to scientific knowledge made by different research groups in the same discipline. This has been developed as part of a study of a small number of major basic-research centres.[1] The main reason for focussing on the centre as the unit of analysis rather than the individual scientist is that over half the annual SRC expenditure goes to support research centres rather than individual scientists or projects.[2] The principal objective of this work is to provide information which may facilitate the taking of decisions of the third and, in particular, the fourth type identified above. In deciding the distribution of resources between research centres, one of the most relevant pieces of information is the past performance of those centres. "Track record" is by no means the *only* factor determining future research performance (there are others like the "ripeness" or potential for exploitation of new work), but it is undeniably one of the most important. If other factors are equal, or alternatively if they are completely indeterminate, then one has to assume that, on average, a new scientific project is *more* likely to be carried out well by a group with a record of successful research over the recent past than by one with a less distinguished record. We are concerned in this paper with the question of the extent to which past performance can be reliably and satisfactorily evaluated. The questions to which we have addressed ourselves take the following form: to what extent can the outputs of research groups or centres working in the same discipline be compared? Is it possible to assess whether one research group has contributed more to scientific knowledge than another, and, if so, how?

The structure of the paper is as follows: the first sections discuss the nature of basic research, and the outputs from it; this is followed by a critical review of the literature on previous uses of various measures of science, drawing out the severe methodological and conceptual problems associated with each measure; on the basis of this, it is shown that, once some consideration is given to the question of what is actually being measured, *careful* use of a number of "partial indicators" of scientific progress can yield information on the relative performance of groups of scientists; the later sections then detail our empirical work in which several partial indicators are combined to assess the performance of a number of large, basic-research centres working in the field of radio astronomy.

2. The nature of basic research and its outputs

While assessing the output from basic research may be extremely difficult, there can be no doubt that there is an output of some kind. This may take the form of new scientific knowledge (theories, empirical findings, and so on), new scientific problems, or new practical ideas or problems (cf. Freeman [14]). In short, there is a flow through basic science of information generally embodied in research publications or in people. Borrowing a model from the conceptual vocabulary of economists, we can usefully conceive of science as an "input–output" process (cf. e.g. Moravcsik [38]). Such a model of science is depicted in fig. 1. While there are few major difficulties here in identifying suitable input measures, there are severe conceptual as well as methodological problems associated with finding appropriate output measures. These stem from the intangible nature of much of the output or "product" of basic research activities. Indeed, the very nature of the "product" depends on our philosophy of science (for instance, on whether we assume scientific knowledge is cumulative in nature, or whether its advance is better represented as a series of revolutionary transformations – cf. Kuhn, [25]), as well as on our approach to the sociology of scientific institutions. Part of the problem is that there are many ways in which contributions to scientific progress are made. While a few scientists are responsible for making major "discoveries", most make relatively small incremental additions to our knowledge (perhaps in the form of more precise measurements). Both types of contribution are obviously essential to scientific progress. In addition, scientists who are primarily teachers, administrators, or technicians, all play crucial roles in scien-

---

[1] The project aimed to assess the performance of five major centres supported by the SRC. These were the Daresbury and Rutherford high-energy physics laboratories, the radio astronomy observatories at Cambridge and Jodrell Bank, and the optical astronomy facilities at the Royal Greenwich Observatory.

[2] The CERN, Rutherford, Daresbury, Appleton, and ILL Grenoble Laboratories, together with the two Royal Observatories, accounted for £87 million of the £137 million spent by the SRC in the financial year 1978–79, while several other major research groups, including the radio astronomy observatories at Cambridge and Manchester Universities, were also supported mainly through consolidated block grants rather than through grants for specific research projects.

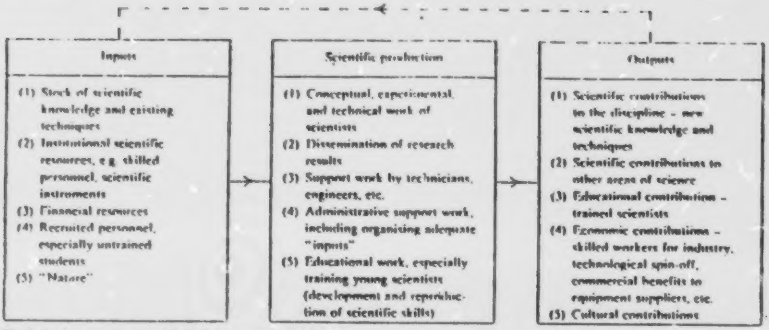

Inputs
(1) Stock of scientific knowledge and existing techniques
(2) Institutional scientific resources, e.g. skilled personnel, scientific instruments
(3) Financial resources
(4) Recruited personnel, especially untrained students
(5) "Nature"
Scientific production
(1) Conceptual, experimental, and technical work of scientists
(2) Dissemination of research results
(3) Support work by technicians, engineers, etc.
(4) Administrative support work, including organising adequate "inputs"
(5) Educational work, especially training young scientists (development and reproduction of scientific skills)
Outputs
(1) Scientific contributions to the discipline – new scientific knowledge and techniques
(2) Scientific contributions to other areas of science
(3) Educational contribution – trained scientists
(4) Economic contributions – skilled workers for industry, technological spin-off, commercial benefits to equipment suppliers, etc.
(5) Cultural contributions

Fig. 1. An input–output model of a basic research discipline.

tific development (cf. Lawani [28, p. 26]). Although there are other such facets to scientific activity, we shall be concentrating here on assessing contributions to scientific knowledge,[3] since these are most directly related to the primary goal of basic research.

There are a number of possible yardsticks for assessing the contributions to scientific knowledge made by individuals or groups of scientists. These include the number of scientific publications produced in a given period or for a given volume of resources, the number of times these publications are cited in other articles or books, the evaluation by scientific peers of the importance of the published work, the number of "discoveries" or other major advances in knowledge, and the recognition afforded to the authors of the publications (in the form of honours or prizes, for example). Some of these measures can be used fairly readily on a small scale, but it is difficult in practice (and perhaps invalid theoretically) to extend such indices of performance across a wide range of disciplines and countries, or over an extended timescale (cf. Freeman [14, p. 4]). It is the contention of this paper that, although no *absolute* quantification of basic research is possible, one can make valid and useful *comparisons* between the scientific performance of different research groups, provided that careful thought is given both to the choice of groups for comparison, and to the question of what the various indicators of scientific performance are actually measuring.

3. Scientific activity, scientific production, and scientific progress

Before looking at the various measures of output from basic research, it is helpful to separate out different aspects of the performance of scientists, and to make a distinction between scientific activity, scientific production, and scientific progress (cf. Moravcsik [37, p. 268]). The first, scientific activity, is concerned with the consumption of input resources, and is therefore related to such factors as the number of scientists involved, the expenditure on their research, the percentage of their time spent on research, and the number of support staff (for example, in the form of technicians or administrators). The second, research production, refers to the extent to which this consumption of resources creates a body of scientific *results*. These results are usually manifest in the form of research communications, although scien-

---

[3] We examine the external benefits from radio astronomy in another paper (Irvine and Martin [22]). This considers both technological spin-off and the migration from radio astronomy of trained students, who take with them skills that subsequently prove useful in a variety of high-technology jobs. However, the main focus of our work has been on contributions to scientific knowledge, because nearly all the members of the research groups whom we interviewed gave this as the primary justification for continued financial support of their research, rather than other outputs such as contributions to education or to technology.

BEST COPY AVAILABLE

tists do communicate through other informal channels such as letters, seminars, and personal conversations.[4] The third, scientific progress, refers to the extent to which scientific activity actually results in substantive contributions to scientific *knowledge* (as judged by other scientists).

The problems of assessing scientific progress stem partly from the fact that these contributions to knowledge are not always regarded as cumulative, which means that such concepts as the "quantity" and "quality" of research (terms apparently implying absolute characteristics) may be misleading. As we shall see below, while some of the output indicators are fairly clearly linked to scientific production, their links with scientific progress are more complex and problematic. Yet it is with indicators of scientific progress that we must be most concerned if we are to evaluate the extent to which scientists succeed in fulfilling the primary goal of basic research, the production of new scientific knowledge.

4. Publications

The first output measure we shall consider is based on scientific publications. Several authors have made use of the distinction between the "quantity" and "quality" of research contributions (e.g. Lawani [28, p. 26]). The former has generally been measured in terms of numbers of publications, and has been the basis for several studies of the growth of science,[5] and of the performance of groups of scientists (e.g. Chang and Dieks [4]).[6] However, such studies have often exhibited a lack of conceptual clarity as to what the number of publications actually measures. While it may be regarded as a reasonable measure of scientific production,[7] its status as an indicator of scientific *progress* is uncertain. Some authors claim to have found a high correlation between numbers of publications and the overall "quality" or "merit" of the papers (e.g. Clark [6]), or the "eminence" of their authors (e.g. Price [47], p. 41]).[8] However, as we shall see below, the relationship of these variables to scientific progress is not always straightforward. Indeed, for some scientists, or groups of scientists, the correlation between "quantity" and "quality" is small or even zero (cf. Smith and Fiedler [50, p. 228]).

The problem is that each publication does not represent an equal contribution to science. Some "mass-producers" of publications make very little scientific progress, while other "perfectionists" achieve in a few publications very significant scientific progress (cf. Cole and Cole [7, p. 382]). Various attempts have been made to circumvent this problem, for example, by "weighting" some publications differently from others,[9] but this has tended to be done without any adequate theoretical basis for the choice of weights (cf. Bayer and Folger [2, p. 382]). Publication counts by themselves fail to "distinguish between the fluency of genius and the loud noises of empty vessels" (*Nature* [43]); they give too much emphasis to the "operator" who produces quantity rather than quality (cf. Bayer and Folger [2, p. 382]. But how can one measure "quality"?

The task of assessing the relative quality of publications has not been made easier by the explosive growth in scientific literature. Over 15 years ago, Maddox [30, p. 15] remarked that, "in so generally flaccid a literature it is difficult to tell which papers are substantial contributions to understanding, and which are but trivial documents", and the situation has certainly not improved since that time. Some discussion has been given to the possibility of weighting publications according to the journal in which they appear (e.g. Garfield [18]). However, there is evidence that even the prestigious journals are not always very discriminating (cf., for example, Lawani [28, p. 26]). Hence, attempts to attach a "quality index" or "impact factor" (Garfield [17, p. 474]) to journals fail to confront the problem of the wide variation in quality *within* each journal (cf. Smith and Fiedler [50, pp. 230–31]). The two principal methods which have been used to obtain an indicator of the quality of publications involve peer evaluation and citation analysis.[10] How these are linked with scientific progress is explored in subsequent sections.

---

[4] These informal channels of communication may be very important, particularly in the initial stages of disseminating new scientific knowledge. However it has been assumed here (although this does need empirical investigation) that much of the information passing through these channels ends up in the final depository of scientific publications, and hence that only research publications need be assessed (cf. Moravcsik [38]).

[5] See Gilbert and Woolgar [20, pp. 279–83] for a summary of some examples.

[6] Numbers of publications are also apparently used by some administrators responsible for allocating resources in science. A recent article in *Nature* [45] refers to a professor of chemistry (and Nobel Laureate) being threatened with expulsion by his university for "inadequate productivity"!

[7] But see note 4.

[8] Price argues that "on the whole there is, whether we like it or not, a reasonably good correlation between the eminence of a scientist and his productivity of papers. It takes persistence and perseverance to be a good scientist and these are frequently reflected in a sustained production of scholarly writing." Similarly, Cole and Cole [7, pp. 387–8] claim that "high producers *tend* to publish the more consequential research….(because) engaging in a line of research is in one sense a 'necessary' condition for the production of high quality work….(and because) the reward system operates in such a way as to encourage the creative scientists to be productive and to divert the energies of the less creative scientists into other channels".

[9] See Bayer and Folger [2, p. 382] for a summary of such attempts.

Without some allowance being made for variations in quality, a simple count of publications, while it may provide a measure of scientific production, is not a *measure* of scientific progress. It is what we shall term a *partial indicator*[11] *of scientific progress* – that is, a variable determined *partly* by (a) the level of scientific progress made by the individual or group, but also influenced by (b) other factors, such as various social and political pressures. These include the publication practices of the employing institution (current *and* previous institutions – see Crane [11, p. 703]), the country, and the research area. Another factor is the emphasis placed on numbers of publications for obtaining promotion, tenure, or grants.[12] The situation is complicated by the fact that these 'other factors' not only vary between scientists or groups of scientists, but may also change with time.[13] Moreover, the question of whether (a) dominates over (b) is problematical. It cannot be *assumed*, a priori, that (b), compared to (a), is relatively insignificant, or even that it comprises a set of entirely random influences whose effects can be taken to cancel out either for all large aggregations of scientists or for an extended time-period. (Some of the effects may be random, but others will be systematic in nature, depending on the particular set of social and political circumstances.) The relative importance of (a) and (b) on publication counts can only be established empirically. In short, we have to consider *why* scientists publish papers, and to realise that they do so not only to present valuable results, but also for social, political, and career reasons.

Treating publication counts as a partial indicator, rather than a measure, of scientific progress helps us to see why it may be dangerous to rely on numbers of publications for assessing the scientific output of an individual, or even of a small group, since it is in these cases that there may be wide, and possibly unpredictable, variations in the relative effects of (b). With larger groups, however, it may be possible to select carefully matched groups for comparison so that we minimize some of the effects of (b) – for example, by selecting groups working in the same specialty, under similar systems of organization and funding, and publishing similar types of papers in the same journals. We *may* then find that the effects of (b) relative to (a) can be reduced sufficiently to make publication counts a useful indicator of scientific progress. Some preliminary empirical results on this are presented in the latter part of this paper.

---

[10] Citation analysis has already found a variety of uses, particularly in the United States, where the National Science Foundation uses it to assess, for example, its funding of chemistry departments; where various universities employ it to decide cases of promotion and tenure; and where citation evidence was produced in court to prove that a woman denied tenure was as good as, or better than, two men promoted over her (Wade [52, p. 429]).

[11] To a certain extent the term "partial indicator" is tautological in that use of the word "indicator" already implies a partial or incomplete measure. However, in view of the tendency of some users of indicators to forget this partiality or incompleteness in the nature of indicators, and the consequent danger of confusing the conceptual with the ontological, i.e. equating the mental construct with "reality", it is worth emphasizing the partiality. Consequently, the term "partial indicator" has generally been used throughout the paper, although where repetition of the word "partial" would be labouring the point unnecessarily, this has been abbreviated to "indicator".

[12] As the use of publications to account for time and money spent on research has grown, so the pressures to publish have increased correspondingly, resulting in a state of "literature inflation" – a vicious circle in which the more papers that are written, the less they count for, and the greater is the pressure to publish more (cf. Margolis [31, p. 1218]).

[13] Cf. e.g. Sullivan et al. [51, pp. 182–84], who find that the productivity of certain theoretical physicists has quadrupled over a 20-year period, while that of experimentalists working in the same specialty has stayed constant.

BEST COPY AVAILABLE

5. Citations

The aim of citation analysis is generally held to be that of injecting a "quality factor" into the evaluation of scientific publications. A fundamental assumption of most previous use of citation analysis is that the impact of a paper lies in its influence on subsequent research papers, and that each instance of such influence will manifest itself by the influenced paper referring to the influencing paper (cf. Moravcsik [38]). On the basis of this somewhat challengeable assumption, the more enthusiastic devotees of citation analysis assert that the number of citations earned by a paper represents the "quality" of that paper (cf. e.g. Cole and Cole [7, p. 379]). [14] Various pieces of evidence have been adduced to support this claim: for example, Clark [6] finds a high correlation between citations and quality, and concludes that a citation count is the best available indicator of the "worth" of research work. Another kind of evidence is said to come from Nobel Laureates, the majority of whom are amongst the top 0.1% most cited authors (cf. Garfield [19, p. 485]), giving rise to the suggestion of "predicting Nobel Prize winners" by citations (Garfield [16, p. 671]). On the basis of this and other evidence, [15] it has been concluded by some that citations are "the best practical indicator of the worth of research" (Porter [46, p. 257]). How justified is this claim?

As most citation analysts are ready to admit, counts of citations face a number of problems, some technical, some more substantive (cf. Cole and Cole [9, p. 24]). Technical problems stem from the fact that the *Science Citation Index*, which is generally used in these studies, does not provide a totally accurate reflection of citation structure (cf. Moravcsik [38]). For example, citations to multi-authored publications are listed only under the first-named author. Since there appears to be a high correlation between citation counts obtained in this way and *total* citation counts (obtained by looking up a list of all the papers by a researcher regardless of whether they were the first-named author), it is generally assumed that using only first-named authors makes little difference (cf. e.g. Cole and Cole [9, pp. 32-34]. While this may be true on average for *most* scientists, it can make a crucial difference to the citation counts of *some* individuals (cf. Lindsey [29]), particularly penalizing those whose work involves a great deal of collaboration. Hence, in comparisons between groups of scientists who engage in differing degrees of collaboration in their research, use of first-named author citations may lead to significant systematic errors, and in the empirical work described below total citations are used for this reason. A second technical problem is that of individuals who are listed under more than one name in the *Science Citation Index*, perhaps because they sometimes use one initial and sometimes two, or alternatively because of a change of name on marrying; these, however, can generally be checked where such variations are suspected. Similarly, some names prove "difficult" for citing authors to spell, and again one may have to check under several different possibilities in the *Index*. Yet another problem is that of two authors with the same name and initials, but this can generally be dealt with provided the research interests of the scientists and the journals in which they customarily publish are known. In the empirical study described later, one difficulty concerned preprints which are listed together in the *Science Citation Index* under "in press". In a small number of cases of collaborative papers it was not possible to establish conclusively whether the cited preprint was the product of the research group under examination, or that of another research centre, but in our study such instances were very rare (less than 0.5% of all citations). Finally, some citing authors make clerical errors in recording page numbers, volume numbers, and journal names. These can generally be identified through comparison with a complete and accurate list of publications for the individual concerned. In short, most of the technical errors can be overcome with care and diligence; and this effort would seem well warranted if Porter's [46, p. 264] estimate that, if uncorrected, these errors could be as large as 25%, is anywhere near accurate.

The substantive problems associated with citation analysis are more complex and severe. First, there is the case of high-quality work that is initially resisted or neglected, perhaps because it is "ahead of its time", and which therefore accu-

---

[14] In a later paper, Cole and Cole [9, pp. 23-24]) make a distinction between "absolute quality" (i.e. papers embodying "absolute truth") and "socially determined quality", but they still maintain that citations "are an adequate measure of the quality of work socially defined".

[15] See e.g. Bayer and Folger [2]; Cole and Cole [7]; Myers [41].

AVAILABLE

mulates few citations. The assertion that such occurrences are rare (cf. e.g. Cole and Cole [9, pp. 24-25]) needs more empirical justification than the quoting of a few famous instances such as Mendel, accompanied by a claim that these occur most infrequently. Secondly, poor quality papers may be frequently cited, because they are either controversial or "mistaken" (cf. e.g. Janke [23, p. 892]). The response of Cole and Cole [10, pp. 32-33] to this criticism is that,

> Much important work turns out in historial retrospect to have been "incorrect"... A paper which is important enough to receive a large number of citations is probably a significant contribution.[16]

While there is more than a little validity to this argument, it illustrates the conceptual confusion surrounding the use of the term "quality", the characteristic which citations are often claimed to measure.

Thirdly, there is a problem in dealing with the small number of extremely important papers (such as those by Einstein) that present new fundamental ideas. Once these ideas have been accepted and fully integrated into the body of scientific knowledge, the original papers may be seldom cited. Cole and Cole [9, p. 31] admit that this may lead to errors for individual papers, but assert, again with little apparent supporting evidence, that this is not a problem when dealing with a number of papers or with scientists. Fourthly, a high-quality paper in a small, narrow, or highly specialized field may produce few citations – fewer than a similar quality paper in a more popular field (cf. McGervey [35, p. 30]). On the basis of comparisons between solid-state physics and the smaller specialty of high-energy physics (although this would appear to be a far from typical "small" field), Cole and Cole [9, pp. 28-29] conclude there is no evidence for such an effect, though they do admit that it might be a problem for very small fields.[17] Other authors dispute this claim by the Coles, arguing that, if citations are used as a measure of quality, one must take account of the size of population of articles that might potentially refer to the work in question (cf. e.g. Sullivan et al. [51, pp. 195-96]) – in other words, the numbers of citations must be "normalized" if one is attempting to compare research in different specialties, or to compare research in the same specialty at different times.

A fifth factor influencing the relationship between numbers of citations and quality is the effect of self-citation, the incidence of which may be partly related to the size or the state of development of the specialty,[18] although it is also structured by a number of other factors including the degree of "openness" of the individual or group concerned to work carried out elsewhere. Matheson's [33, p. 208] study of British university chemistry departments reveals that self-citation rates can vary considerably (by a factor of two) between different research groups working in the same field, and this suggests that, at least for groups practising a high degree of self-citation, the number of citations received may be at least partially determined by the number of publications they produce. In consequence, numbers of citations may reflect "quantity" as well as "quality" of publications.

Another problem with the use of citations as a measure of quality is that certain *kinds* of papers are more frequently cited than others of similar quality. For example, authors in theoretical physics

---

[16] The Coles [10, pp. 32-33] continue: "Why should a large number of scientists waste their time pointing out a trivial error? In fact they do not. Papers which are trivial and receive critical citations will not accumulate large numbers of citations."

[17] Citation patterns within a specialty *do* change considerably with the size of the specialty. For example, Meadows and O'Connor [34] show that the average number of references per paper jumped from 7.1 to 9.9 during the first year of research in the new and rapidly growing field of pulsar research. Similarly, Sullivan et al. [51] note a doubling in the average number of references per paper over a 20-year period in the research area of weak interactions, a sub-field of high-energy physics. In contrast, it is argued by some (e.g. Lawani [28, p. 30]) that the size of the research field will not greatly influence the relationship between the quality of a publication and the number of citations it receives because, although there are fewer papers to cite from in a small field, each paper is also more prominent than one of similar quality in a broader field. However, there must be doubts as to whether these two opposite size-effects do exactly cancel out, and hence whether the relationship between quality and citation does "scale up" exactly in this way.

[18] For example, Meadows and O'Connor [34, p. 97] find that the percentage of self-citations dropped from 15% of the total citations to 10% over the first year of pulsar research as more, potentially citable work by other authors became available.

BEST COPY AVAILABLE

frequently refer to accounts in the mathematical literature of the development of a new technique simply as a means of shortening their own exposition. Similarly, in chemistry and biology, papers describing a standard technique [19] may be highly cited regardless of whether or not the first published description of the technique constitutes a high-quality piece of research (cf. *Nature* [44, p. 699]). Hence, the number of citations a paper receives may reflect its nature as well as its quality.

Finally, the use of citation counts is complicated by the "halo effect" – the fact that, when there is a choice of publications to cite, the author of a paper is more likely to refer to the work of eminent scientists (cf. Smith and Fiedler [50, p. 229]). Scientists who have gained a reputation by publishing significant research in the past then enjoy a halo effect in that all of their research, even if some of it is of lower quality, gains additional attention. Cole and Cole [9] admit that, if one considers the total number of a scientist's citations, some will be due to the halo effect, but they argue that the size of the effect should be directly related to the significance of the scientist's research and that this does not therefore affect the use of citations as a measure of quality. There are, however, at least two major weaknesses in this argument. First, the quality of currently cited work may not be as good as that of the previous research that determines the size of the halo effect (cf. Croom [12, p. 1173]). Secondly, the relationship between the halo effect and the quality of past research may not be a *linear* one, with the result that analysis based on numbers of citations could *exaggerate* the differences between high quality and low quality research. [20] For example, the halo effect enjoyed by a Nobel Prize winner is probably far larger than that accorded to the winner of a British Royal Society medal (often for work of not greatly inferior quality) because of the tremendous publicity given by the semi-scientific and popular press to the former. [21]

Underlying all these problems with the use of citations as a measure of quality is our ignorance of the reasons why authors cite particular pieces of work and not others. The problems described above arise principally because simple citation analysis presupposes a highly rational model of reference-giving, in which citations are held to reflect primarily scientific appreciation of previous work of high quality or importance, and potential citers all have the same chance to cite particular papers (regardless of who the authors are, where they published, what language they wrote in, etc. – in other words a "free market" in scientific ideas). The model also assumes that the norms of citation behaviour are largely unaffected by any "external" pressures (including the awareness of scientists that citation analyses may be used to assess their scientific performance), and that all citations are of equal value or intent. Such a model is obviously a grossly oversimplified and possibly highly misleading representation of reference-giving; the activity of reference-giving is, like other aspects of scientific work, a social activity, and an author's citation practices can only be fully understood if they are related to individual cognitive interests, institutional position, and social and political goals. Before we look to see how this can be done in practice, certain conceptual distinctions must be made in order to ascertain what the numbers of citations may or may not indicate.

6. The quality, importance, and impact of publications

Previous use of citation measures has, as we have seen, mostly exhibited a lack of conceptual clarity as to what the citation rates measure. To overcome these problems, some authors have attempted to distinguish between the "quality" and the "impact" of publications, arguing that, even if citations do not provide an indicator of the quality of a paper, they do at least reflect its impact on the scientific community (e.g. Garfield [15, p. 30]).

---

[19] O.H. Lowry, the most highly cited scientist, regularly receives several thousand citations each year (several times more than even the most eminent of scientists), the bulk of which refer to a technique for protein determination. Given that this technique is now so standard, one wonders whether all those who cite this work have in fact read the relevant paper before they make use of it.

[20] This may partly account for the controversial conclusion of Cole and Cole [8, p. 368] that "only a few scientists contribute to scientific progress".

[21] Cf. Inhaber and Przednowek [21, p. 33]: "In practice, the semi-scientific and popular press considerably distort the visibility of the different forms of recognition. This may produce a gap between the perceived and true quality of scientific work."

BEST COPY AVAILABLE

However, few have specified exactly to what each of these concepts refers. We would argue that it is necessary to distinguish between, not two, but three concepts – the "quality", "importance", and "impact" of the research described in a paper – if we are to understand what, if anything, citation-rates measure (cf. Lawani [28, p. 27]). The first of these concepts refers to the research itself, while the latter two are more external, referring to the relations between the research and other research areas, and describing the strength of the links to, or the implications for, other research activities.

"Quality" is a property of the publication and the research described in it. It describes how well the research has been done, whether it is free from obvious "error", how aesthetically pleasing the mathematical formulations are, how original the conclusions are, and so on. But quality is still relative rather than absolute, and it is socially as well as cognitively determined; it is not just intrinsic to the research, but is something judged by others who, with differing research interests and social and political goals (i.e. different cognitive and social "locations" – cf. Martin [32]), may not place the same estimates on the quality of a given paper. Even the same individual may evaluate the quality of a paper differently at different times because of progress in scientific knowledge and shifts in his or her location.

The "importance" of a publication refers to its *potential* influence on surrounding research activities – that is, the influence on the advance of scientific knowledge it would have if there were perfect communication in science (in short, if there were the completely "free market" of scientific ideas mentioned above). However, there are "imperfections" in the scientific communications system, the result of which is that the *importance* of a paper may not be identical with its *impact*.

The "impact" of a publication describes its *actual* influence on surrounding research activities at a given time. While this will depend partly on its importance, it may also be affected by such factors as the location of the author, and the prestige, language, and availability, of the publishing journal (cf. Dieks and Chang [13, p. 247]).

Having distinguished these three concepts, we must next examine how they are related to each other, and to the notion of scientific progress described earlier. The importance and the impact of a publication are determined by the relations between the research described in it and the surrounding specialty, and hence by the quality of the former and by certain characteristics of the latter – for example, the level of research activity it exhibits. This level of activity in turn reflects scientists' perceptions as to which specialties deal with the most "fundamental" questions (Cole and Cole [10, p. 32]), which find it easiest to attract research funding, which are more likely to attract high quality students, and so on – in other words, factors dealing with the relationship between the narrow research area and the wider research domain. It seems plausible to suggest that a high quality paper in an active specialty will generally have a greater importance than a similar quality paper in a dormant or declining one.[12] If so, we can see that the "quality" of a publication need not be synonymous with its contribution to scientific knowledge; for example, a high-quality paper in a stagnating field may contribute little to the advance of scientific knowledge in general. Nor is the "importance" of a paper necessarily an indication of the size of its contribution; important papers can go unnoticed if authors express themselves poorly, if they publish in a journal with a restricted circulation,[13] or if they have not previously been particularly prominent in the scientific community.[14] It is the "impact" of a publication that is most closely linked to the notion of scientific progress – a paper creating a great impact represents a major contribution to knowledge *at that time* (although its impact may of course alter with time).

Is it possible to obtain any absolute or direct measure of the quality, importance, or impact of a publication? The short answer is "No". These factors are not absolute (in the sense of holding the same value for all people and all time) but relative, varying over time and according to the

---

[12] Unless the work is so exceptional that it revitalises a previously stagnating research area, sending reverberations far beyond its boundaries to neighbouring areas.

[13] Cf. Lawani [28, p. 31], who notes that entomological articles published in West Africa are cited less than those published overseas.

[14] Cf. Whitley [55, p. 230], who finds some evidence for this "bandwagon" or "halo" effect: "once a paper by a man is cited there may be some incentive to cite other papers by him. Partly it may be due to the intrinsic quality of the papers and partly to the high visibility of the man and his work."

BEST COPY AVAILABLE

cognitive and social location of the assessor. Moreover, they cannot be evaluated directly, but only apprehended indirectly through the perceptions of other scientists (by peer evaluation), or partially inferred from the social practices (for example, citation practices) of scientists. The number of citations to a publication is not a direct reflection (a "measure") of its quality or importance, nor even of its impact. The citation rate is a *partial indicator* of the impact of a scientific publication: that is, a variable determined partly by (a) the impact of the paper on the advance of scientific knowledge, but also influenced by (b) other factors, including various social and political pressures such as the communication practices (for example, the reading and referencing habits of different individuals in different institutions, countries, and research areas), the emphasis placed on numbers of citations for obtaining promotion, tenure, or grants,[25] and the existing visibility of authors, their previous work, and their employing institution. As with numbers of publications, it cannot be assumed that the effects of (b) are relatively insignificant compared to those of (a),[26] nor that (b) comprises a set of essentially random influences, the effects of which cancel out in analyses of large aggregations of scientists or for extended time periods. The "other factors" that make up (b) are largely social and political rather than purely "scientific", and, while some of their effects on citation rates may be random, others can be expected to vary in a systematic way between individual, or groups of, scientists occupying different cognitive and social locations. The relative importance of (a) and (b) on citation rates can only be established empirically – for example by looking to see whether there are systematic variations in referencing behaviour between different groups of scientists.[27]

Once the claim for citation rates is restricted to that of being a *partial* indicator of the scientific *impact* (rather than a *measure* of the *quality* or importance) of research publications, then some of the methodological problems associated with citations that have been detailed above are considerably diminished. First, an important pioneering paper may be little cited initially, but this would suggest that such a publication also has little *impact* at first until its importance is eventually recognised, so the number of citations is still correlated with impact. The small number of early citations is therefore a reflection of the structure and organization of the scientific community and its communication system, rather than a consequence of any inherent weakness in citation analysis per se (cf. Lawani [28, p. 29]). Indeed, the use of citation rates to identify those pioneering papers whose importance is not immediately recognised may well be an essential first step in ascertaining *why* recognition of their importance is not always readily forthcoming. Secondly, a controversial, but low-quality paper, or even a "mistaken" paper, may create a large impact, stimulating in its wake further research aimed at improving upon, or refuting it. If such a paper generates a large number of citations, this can be regarded as a reflection of its impact, rather than of its quality (cf. Cole and Cole [9, p. 25]), and hence of how much it stimulates and therefore contributes to scientific progress. Thirdly, if citation rates are taken to reflect the impact of a paper rather than its quality or importance, the fact that, in cases where there has been complete integration of certain basic ideas, the original publications currently receive few, if any, citations, can be interpreted as implying that

---

[25] Kaplan [24, p. 183] warned that the use of citation analysis as a tool for evaluating science could lead to changes in citation practices, and in the course of our interviews with astronomers, several mentioned their belief that some individuals and groups, particularly in the United States, had already begun to exploit the system through informal mutual-citing arrangements – often termed "citation circles".

[26] Although they discuss the possible effects of various "distorting" factors on citation counts, Cole and Cole [8, p. 369] assume that the incidence of such effects is rare and their magnitude insignificant, thus enabling them to claim that "citations generally represent an authentic indicator of influence". (It should be noted that this represents a weakening from their earlier claim that the number of citations represents the *quality* of a paper – see Cole and Cole [7, p. 379].)

[27] Kaplan [24, pp. 182–3] raises some of the questions that need to be addressed in such empirical studies. For example, "what is the influence of the organisational context and of one's colleagues? Are there informal norms about citing the works of colleagues, especially of superiors, so long as there is some remote connection between their work and the work being reported?... Do citation practices reflect significant elements of the normative and value systems of scientists?" These questions have largely gone unanswered since then, although some recent work on the classification of citations by context has been carried out by Chubin and Moitra [3], and Moravcsik and Murugesan (e.g. [39,40]) – work that may eventually provide answers to such questions.

BEST COPY AVAILABLE

those original publications have little direct impact now. For example, Einstein's 1905 papers are still regarded as being of supreme importance, yet they have a low impact today and elicit few citations because hardly anyone actually reads the papers. Before they reached such a stage of integration into scientific knowledge, these papers had a far higher impact and would presumably have received a very high number of citations compared with other contemporary publications. Since that time the impact of these papers has diminished, while the importance of the ideas contained in them has remained high and has been transferred through second, third, and successive generations of papers, each possibly citing the "parent" generation of papers but not the "grandparent" and previous antecedent papers (cf. Margolis [31, p. 1215]). Hence, there seems to be no objection to using citation rates as a partial indicator of "impact" provided that we recognise that this is not synonymous with "importance", a quality which may be transferred through successive generations of papers in such a way that citations do not reveal the "founding father" paper.

The fourth problem, that of the high-quality paper in a small unpopular field which gains less citations than those in large active fields, has already been dealt with – the lower number of citations being interpreted as representing a smaller impact on the advance of scientific knowledge (in the sense of having implications for the work of a smaller number of researchers). Similarly, the fact that certain kinds of research papers (or papers written by more "visible" scientists) tend to receive more citations than average can be interpreted as indicating that such papers, whatever their quality, do create a wider impact than average. For example, a paper describing a minor improvement to an established method may constitute a relatively small or low-level contribution to scientific knowledge, but when integrated with other contributions it can have a big impact on scientific activities (and hence on subsequent scientific progress), altering the procedure involved in some standard and widely used technique; and this will probably be reflected in its citations, at least until it is integrated into the body of scientific knowledge.[28]

The problem of self-citation is a little more complex in that it is not immediately clear whether, when using citation rates as a partial indicator of impact, one needs to exclude self-citations or not. It can, for example, be argued that individuals or groups of scientists who engage in a high level of self-citation do so because their previous work has had a large impact on their current research. However, at the same time, it must be recognised that in the case of self-citations, the relative effect of (b), the "other factors", on the citation rate may very well be greater than in the case of "normal" citations. Perhaps the best solution is to examine empirically the effect of including both self-citations and in-house citations (citations to the work of colleagues within the centre) in order to see whether the effect is significant or not, and this is attempted in the empirical study described below.

7. Peer evaluation

The method of assessing contributions to scientific knowledge apparently most favoured by scientists is that of peer evaluation.[29] However, as with publication counts and citation analysis, peer evaluation does not yield a simple measure of scientific progress. The difficulty centres on the subjective nature of this method: peer evaluation is based on individual scientists' *perceptions* of contributions by others to scientific progress, perceptions arrived at through a complicated series of intellectual and social processes, mediated by factors other than the quality, importance, or impact

---

[28] There is, however, a problem with citations in that some standard techniques, even after having become fully accepted, still do generate citations to the original papers (as in the protein-determination techniques referred to in note 19, while others (for example, aperture-synthesis in radio astronomy) do not. The reasons for such differences in citation patterns need further empirical investigation.

[29] Out of the 69 scientists interviewed in this study, 76% saw peer evaluation as the best method, and 19% saw it as equal best with either publication counts or citation analysis. Fairly similar results are obtained by Chan [3] in his study of university faculty staff and administrators on the relative importance of various output indicators in evaluating the effectiveness of academic research. Peer evaluation and "articles published in prestigious journals" are seen as the most valid indicators, although it is not absolutely clear in this study whether the latter refers merely to numbers of publications, or whether it was interpreted by the respondents in the survey as including some assessment of the quality or influence of these papers as well.

BEST COPY AVAILABLE

of the research under evaluation. This gives rise to three main sets of problems in using peer evaluation, and we need to be aware of these if we are to use it as another partial indicator of scientific progress.

First, and most importantly, we must recognise the effect of political and social pressures within the scientific community on the way certain scientists will evaluate the worth of their peers. Modern science is highly competitive, with groups vying for prestige and funding. Research groups, particularly those in areas of science characterized by rapid scientific growth, complete fiercely for the recognition and acceptance of their ideas and theories, and peer review can reflect the pressures associated with this rivalry. Scientists who are asked to evaluate research may be influenced by the possible implications of the peer evaluation both for their own futures and for those of their colleagues. Of course, one could try to reduce the effects of this problem by asking every peer even remotely affected by the particular scientific contribution under assessment for his or her evaluation, and then aggregating or averaging the responses; but given that this is seldom practical, it is generally only possible to ask a representative selection of peers – and this in turn raises all those statistical problems associated with sampling. One can at best only attempt to ensure that representatives from all the important groups in the field are involved in the assessment in order to minimise these problems.

Representative sampling also helps reduce the effects of a second problem with peer evaluation, namely that of allowing for the diversity in the cognitive locations of the reviewers.[30] There is inevitably a tendency for scientists to evaluate the worth of a scientific contribution in terms of their own research interests and activities. This problem is compounded by the fact that, although peer evaluation is supposedly based on the published output of scientists, some reviewers have not read all the relevant publications, with the result that their evaluations may be based more on informal "coffee-room" discussions or on the *reputation* of scientists (in particular, their reputation for performing at conferences) than on their actual contributions to scientific progress (cf. Smith and Fiedler [50, p. 226]). One way to overcome this problem is to include in the sample of reviewers only eminent scientists with a wide knowledge of the field concerned, but this can introduce other problems (such as that of excluding the views of less well established, but often no less relevant, schools of scientific thought).

Finally, we should mention a problem associated with all attempts to gauge opinion – the propensity of people to conform to conventionally accepted patterns of belief. Although perhaps privately holding the work of certain other well-known scientists in low esteem, they may publicly express different views (even if they know that their assessments are to be treated confidentially). Such conformist behaviour need not always be the result of a conscious decision, one example being the "halo effect" by which work may be evaluated more highly because of its association with a successful group or a prestigious university. Furthermore, particularly in those cases where evaluators do not possess all the knowledge needed to make a balanced judgement, yet do not want to withdraw because an invitation to participate in peer evaluation is often regarded as an indication of status within the scientific community, there can be a tendency to base the evaluations on the *recognition* accorded to work, for example in the form of prizes or medals, rather than on the work itself. There is then the additional problem of taking account of the time-lag between the making of the contribution to science, and its subsequent recognition.

Because of these difficulties, peer evaluation cannot claim to provide an incontrovertible measure of scientific progress. As with the other measures we have considered, it is at best a partial indicator of scientific progress[31] – a variable influenced partly by the size of the contribution to scientific progress and partly by other factors, the relative importance of which cannot be assumed to be insignificant.

---

[30] Cf. the discussion by Mitroff and Chubin [36, p. 199] of the need to examine the cognitive styles of reviewers and reviewees in order to make sense of the results of peer review.

[31] Hence, one would not expect the results of peer evaluations to be altogether independent of other quantitative indicators such as publications and citations. The former is linked to the latter in that peer evaluation is based partly on journal literature, and some aspects of journal literature (for example, the acceptance of papers for publication, and the citing of others in a paper) may in turn be partly determined by peer evaluation (cf. Moravcsik [38]). In the few studies where peer evaluation is combined with other indicators of scientific progress, the various methods of evaluation are found to coincide fairly well (cf. *ibid.*).

8. Other possible indicators of scientific progress

Two further indicators of scientific progress are the number of "discoveries" or other major contributions to the advance of knowledge, and the formal recognition afforded to scientists by the scientific community. In the case of "discoveries", the crucial questions centre on *who* should identify the "discoveries", and using *what criteria*. The use of the notion of "discoveries" therefore involves problems similar to those encountered with peer evaluation. It would be most unsatisfactory, for example, merely to use text-books or histories of science for identifying discoveries, since one would be relying on the authors' unspecified criteria for judging which contributions to knowledge are sufficiently influential to merit mention as "discoveries", and which are not. Moreover, such judgements reflect *prevailing* perceptions of influence, which can be substantially different from those held at the time the particular contribution to knowledge was made. A contribution that is initially highly influential may subsequently come to be judged as "incorrect" and be omitted from reconstructions found in "official histories", even though it may have stimulated much important research at the time. Alternatively, one might use review articles for identifying discoveries on the grounds that, because the time-delay is shorter, the problem of the reconstruction of history should be less severe. Such a procedure would in effect involve looking at the references cited in review articles, thus bringing one face-to-face with the problems of citations mentioned earlier. It is true that the citation practices in review articles *may* be less wayward or self-centred, and less subject to the effects of the "other factors" described above, than citation practices in research publications (perhaps because a more comprehensive literature-search generally precedes the writing of a review, or because only the "better" scientists are chosen to perform this task). However, this can only be investigated empirically – it cannot be assumed.

The other indicator of scientific progress we should consider is the recognition accorded to scientists through the awarding of various honours such as medals, prizes, invitations to lecture at major conferences, or election to a national academy of science. The allocation of such recognition involves a prior process of peer evaluation, and is, as such, based not only on scientists' perceptions of contributions to knowledge, but also on numerous other factors, the effects of which may vary considerably. For example, Crane [11, p. 710] claims to find that affiliation to a prestigious university is more likely to lead to recognition for a scientist than high productivity. Such a result leads to doubts about the adequacy of recognition as even a partial indicator of contributions to scientific progress, unless preceded by a thorough analysis of the social structure of the scientific community and of its mechanisms for allocating recognition.

9. A methodology for assessing contributions to scientific knowledge

In preceding sections we have seen how all quantitative "measures of science" are at best partial indicators, influenced by a network of interrelated factors of which the size of contribution to scientific progress is but one. Nevertheless, given that all decisions on the allocation of resources between different areas of science must involve some assessment of the respective merits of the various areas, we would argue that selective and careful use of these indicators (despite the associated methodological and conceptual problems) is far better than none at all (cf. Moravcsik [38]). Previous use of these indicators, in the field that has come to be known as "scientometrics", has met with a barrage of criticism, in particular from scientists. The purpose of the previous sections has been to show why much of this criticism is justified. For example, assessments of contributions to science based on numbers of citations can be misleading or meaningless when little or no thought has been given to the question of what the numbers actually indicate. Nonetheless, the fact that indicators of science have in the past been applied in a simplistic or irresponsible manner does not mean that they cannot be usefully employed at all. On the contrary, having analysed the shortcomings and conceptual difficulties associated with each of them, we are now in a better position to see how they might be used to achieve more reliable assessments of contributions to scientific knowledge.

Given the absence of any adequate single measure of scientific progress, the only way forward would appear to be through the combination of *several* partial indicators. However, since each partial indicator is influenced to a greater or lesser extent by numerous other factors in addition to the magnitude of the contribution to science involved, one must attempt to "control for" these other factors by choosing the groups of scientists to be compared in such a way that one can examine the size of the effects of the "other factors" on each of the partial indicators. Only in those cases where convergent results are obtained can it be assumed that the influence of the other factors has been kept small (i.e. the matching of the groups has been largely successful), and that the indicators therefore provide a reasonable estimate of the contribution to scientific progress made by different research groups. The extent to which convergent results can be obtained is examined in the empirical study described here.

As mentioned before, the study focusses not on individual scientists [12] but on research centres, because these absorb a significant proportion of basic research funds. With this unit of analysis, several of the problems with indicators discussed earlier are diminished. In particular, the influence on the indicators of the size of the contribution to scientific knowledge is less likely to be swamped by the influence of those "other factors" whose effects are statistically random, while it is also easier to match groups than individuals [13] in such a way that one controls for some of the "other factors" whose effects are systematic. Fig. 2 summarizes the main problems with the various indicators and how we have attempted to minimize their effects in our study.

In selecting research centres for comparison, what is ideally required is that there should be two or more groups, working in the same specialty over a similar time-period, publishing in the same journals, supported with a roughly similar level of resources, and situated in a similar institutional context. For example, a high-energy physics group like that using the Rutherford Laboratory might be compared with that for Daresbury Laboratory (and both with similar centres overseas). Similarly, the radio astronomy groups at Cambridge, Jodrell Bank, and a number of overseas institutions can be compared in terms of their contributions to scientific progress. It is with radio astronomy that this study is concerned.

Even when dealing with groups of scientists rather than individuals, however, no "matching" is perfect. Thus, while Cambridge and Jodrell Bank are both engaged in radio astronomy, they tend to concentrate on different research areas and problems within this specialty. This is partly the consequence of the different nature of the research equipment at each centre; Cambridge has concentrated on interferometers (for short-baseline interferometry), while Jodrell Bank has placed more emphasis on "big dishes" (used either individually, or together for long-baseline interferometry). It is therefore essential to include other centres in this study to ascertain whether this divergence in emphasis leads to systematic differences in the effects on the partial indicators. For this purpose, the Netherlands Foundation for Radio Astronomy (NFRA) group based on the Westerbork interferometer and the receiver at Dwingeloo, and the Max-Planck-Institut für Radioastronomie (MPI) group based on the Bonn "big dish", were chosen. [14] These latter two centres are essentially national facilities (as opposed to the university facilities at Cambridge and Jodrell Bank), and are both supported at several times the level of their British counterparts. Nevertheless, given the similarities in the size and scale of activity of the

---

[12] The great majority of previous attempts to assess contributions to scientific knowledge have concentrated on individuals rather than research groups, even though very little scientific research is now carried out by isolated individuals. The few authors who have compared groups of scientists include Westbrook [54], Larabi [26,27], Matheson [33], and Andrews [1].

[13] In the case of individuals, it is unlikely that one will find more than a very small number whose (fairly narrow) research interests coincide or even overlap reasonably completely; with groups, however, a wider range of research interests is covered and it should be possible to find several for which there is a substantial degree of overlap.

[14] Inevitably, this matching of centres is still not perfect. The Bonn big dish is more precise than the one at Jodrell Bank, and is therefore used for work at shorter wavelengths and for more spectroscopy. In the case of the Dutch and the two main Cambridge interferometers, the latter are used primarily for continuum work, while the former carries out both continuum and spectral-line work. Nonetheless, the degree of comparability between the centres, while not complete, is sufficient for the analysis that follows.

AVAILABLE

Partial indicator based on	Problem	How effects may be minimized	
(A) Publication counts	(1) Each publication does not make an equal contribution to scientific knowledge	Use citations to indicate average impact of a group's publications, and to identify very highly cited papers	Use only indicators that yield convergent results
	(2) Variation of publication rates with specialty and institutional context	Choose matched groups producing similar types of papers within a single specialty	
(B) Citation analysis	(1) Technical limitations with *Science Citation Index*:		
	(a) first author only listed	Not a problem for research groups	
	(b) variations in names (c) authors with identical names (d) clerical errors	Check manually	
	(e) incomplete coverage of journals	Not a serious problem for "Big Science"	
	(2) Variation of citation rate during lifetime of a paper – unrecognised advances on the one hand, and integration of basic ideas on the other (3) Critical citations (4) "Halo effect" citations	Not a problem if citations are regarded as an indicator of impact, rather than quality or importance	
	(5) Variation of citation rate with type of paper and specialty	Choose matched groups producing similar types of papers within a single specialty	
	(6) Self-citation and "in-house" citation (SC and IHC)	Check empirically and adjust results if the incidence of SC or IHC varies between groups	
(C) Peer evaluation	(1) Perceived implication of results for own centre and competitors may affect evaluation (2) Individuals evaluate scientific contributions in relation to their own (very different) cognitive and social locations (3) "Conformist" assessments (e.g. "halo effect") accentuated by lack of knowledge on contributions of different centres	(1) Use a complete sample, or a large representative sample (25% or more) (2) Use verbal rather than written survey so can press evaluator if a divergence between expressed opinions and actual views is suspected (3) Assure evaluators of confidentiality (4) Check for systematic variations between different groups of evaluators	

Fig. 1. Main problems with the various partial indicators of scientific progress and details of how their effects may be minimized.

Dutch and German groups, their operation as national facilities, and their use of the same journals for publication, it should be possible to see whether carrying out research on an interferomete rather than a big dish leads to systematic dif ferences in numbers of publications or citation

Table I
Number of astronomy researchers in 1978

	Cambridge	Jodrell Bank	NFRA	MPI
Number of astronomy researchers [a] (excluding students)	17	25	~ 40	~ 60
Number of research students [b]	16	14	~ 10	~ 15
Total number of astronomy researchers (staff and students)	33	39	~ 50	~ 75

[a] This is the number of astronomers who regularly use the observation facilities of the group and who publish in astronomical journals. It excludes short-term visiting users of the radio telescopes, but it does include a number of scientists in each group whose primary duties involve the development or maintenance of instrumental and computer facilities, but who nevertheless devote a significant amount of their time to astronomical research.

[b] This excludes M.Sc. students at Jodrell Bank, first-year postgraduates at Cambridge, "Doctorandus" students in Holland, and Diploma students in Germany; in other words, we are comparing like with like, as far as possible.

rates. If there are such systematic differences, and if they are consistent with the results based on peer evaluation, then one can perhaps conclude that these reflect actual differences in the level of scientific progress achieved by each centre rather than merely being a consequence of the "other factors" affecting the publication and citation indicators.

The main partial indicators of scientific progress used in this study are publications, citations, and peer evaluation. "Recognition" is not employed because most prizes and honours are predominantly national,[35] and so, although it could be used for comparisons between Cambridge and Jodrell Bank, it is not appropriate for comparisons between centres in different countries. There are few truly international honours, the Nobel Prize being almost the only exception, and the numbers are certainly too small for systematic use in a comparison such as this.[36] There are similar problems with an indicator based on the notion of "discovery". The number of major discoveries is rather small for this to be used in a systematic way, while broadening the category of "discoveries" may involve some arbitrariness as to what should be included as a "discovery", and what should not.[37] However, some attempt to assess relative numbers of discoveries is made by adopting highly cited papers as an indicator.

10. Input measures

Before considering each of the partial indicators of scientific progress, we must first look at the scale of scientific activity at each centre; there are appreciable differences in the sizes of the four groups, and it would therefore be misleading to compare the scientific progress made by the smallest group with that made by the largest without taking account of their respective scales of activity. There are a number of input measures that can be used to define this scale of activity, one of the most important being the number of scientists actually engaged in research. The 1978 figures[38] for this are set out in table 1. It can be seen that, in terms of researchers (excluding students), Jodrell Bank is approximately 50% larger than Cambridge, the Dutch group over twice as large, and the German group over three times the size. If research students are included, however, these differences are to some extent reduced.

---

[35] For example, it is doubtful whether one can equate a Fellowship of the Royal Society in Britain with membership of the National Academy of Sciences in the United States because of differences in the percentages of practising scientists selected in each country, in the criteria used in selection, and so on.

[36] It is probably significant, however, that the Cambridge group, which, as we shall see later, appears to have contributed most to scientific progress, includes two scientists who shared the Nobel Prize for Physics in 1974.

[37] This is reflected in the responses made by astronomers to a question on what have been the major achievements of each of the groups. There is, for example, a high degree of consensus that the first observations of pulsars represented a major discovery, but far less consensus over various lower-level achievements.

[38] Ideally, the corresponding figures for earlier years should also be included, but these have proved far harder to obtain. Nonetheless, one can make certain general comments about the trends over time. At the start of the ten-year period upon which we have focussed, the Dutch and German groups were in their infancy. They grew rapidly over the first five years or so, but over the next five years (up to 1978), the increases in staff numbers and running costs have been much slower, approaching the modest growth rates experienced by the British groups over the ten years. Hence, during those last five years, the *ratios* of the scale of activity at the four centres have not changed very significantly.

A second factor influencing the level of scientific activity is the amount of time radio astronomers must devote to teaching activities. At both Cambridge and Jodrell Bank, faculty staff report that they spend on average about 30% of their time on teaching and administration of undergraduate and M.Sc courses, while Dutch and particularly German astronomers have far lighter teaching loads. Clearly the more teaching that is carried out, the less time that remains for research, and one way of considering this is as a reduction in the effective number of astronomy researchers using each of the centres. The respective adjustments for each centre are shown in table 2. It can be seen that this accentuates the differences in scale of research between the two British centres and their two continental counterparts.[39]

Table 2

Effective number of astronomy researchers in 1978 after allowing for teaching duties[a]

	Cambridge	Jodrell Bank	NFRA	MPI
Numbers of researchers involved in teaching	14	17	~ 30	~ 20
Percentage of time spent on teaching[b]	33%	31%	~ 10%	~ 10%
Number of researchers effectively "lost" to research	5	5	~ 3	~ 2
Effective number of researchers (excluding research students)	12	20	~ 37	~ 58
Effective number of researchers (including research students)	28	34	~ 47	~ 73

[a] This covers time spent on teaching undergraduate, M.Sc., Dutch "Doctorandus", and German Diploma students.

[b] Individual astronomers were asked how much of their time was devoted to teaching, and the results were averaged for each centre to give these figures.

A third factor to be considered is the annual running costs[40] for the group. The 1978 values for these are presented in table 3, together with various figures on the costs per researcher depending on whether students are included or not, and on whether differences in teaching duties are allowed for. The figures on total running costs show similar or slightly larger differences in scale than the figures in tables 1 and 2, the larger differences being attributable to the higher salaries in Holland and Germany (see the figures on costs per researcher). In addition, the major capital investment costs (that is, the cost of a new telescope, or of extensive refurbishing of an existing instrument) are summarized in table 4. It should be noted that, if these are spread over the likely life-time of a radio telescope (20 years, for example), they are small in comparison with the running costs, even after allowing for the effects of inflation. Consequently, it is the running costs that largely determine the ratios of the total costs for the centres.

A final factor that affects the level of scientific activity in a research group is the number of support staff (i.e. non-astronomers), and the figures for this are shown in the second row of table 5, along with values for the ratio of support staff to astronomers. It can be seen that Cambridge has the smallest number of support staff, Jodrell Bank about 15% more, the Dutch group double the Cambridge number, and the German group is nearly three times better off in this respect. However, the ratios of support staff to astronomers are broadly similar for all the centres.

To summarize: all the input indicators suggest that the scale of scientific activity is greatest for the German group, which is well over twice as large as the smallest group, at Cambridge. Jodrell

---

[39] Astronomers also spend a certain part of their time in research-related administration both within their centre and outside (on national and international committees, etc.). In addition, some researchers devote much effort to developing new equipment. The time taken up with such duties has not been separated from that devoted to research, for the reason that these duties are an essential and integral part of modern, large-scale research. While research centres overall spend a roughly similar fraction of their time on instrumental tasks and administration, this fraction obviously varies cyclically as a centre passes through the stages of designing a major research facility, construction, exploitation, designing a new facility, and so on. As a result, in the 10-year period under consideration, Jodrell Bank claim that they devoted rather more of their efforts to instrument-building than their rivals.

[40] Because the *total* expenditure at a centre can obviously fluctuate appreciably from one year to the next depending on whether a major new piece of equipment is being built, in calculating the running costs for each centre, we have excluded major items of capital expenditure (of greater than £100,000). However, 'running costs', as defined here, do cover *all* other expenditure, including the cost of purchasing less expensive and more routine items of equipment.

Table 3
Annual running costs for 1978 [a] (£ Sterling)

	Cambridge	Jodrell Bank	NFRA	MPI
Annual running costs [b]	£0.6 M	£0.9 M	£2.6 M	£3.7 M [d]
Annual cost per astronomy researcher (excluding students)	£35 k	£35 k	£65 k	£60 k
Annual cost per astronomy researcher (including students)	£20 k	£25 k	£50 k	£50 k
Annual cost per effective researcher [e] (excluding students)	£50 k	£45 k	£70 k	£65 k
Annual cost per effective researcher (including students)	£20 k	£25 k	£35 k	£50 k

[a] This is the estimated expenditure in the calendar year 1978 after excluding the major items of capital expenditure shown in table 4. Where the financial year does not coincide with the calendar year, an appropriate average of the expenditure in financial years 1977/78 and 1978/79 has been taken. Because of this, and because of the different accounting procedures at each of the centres, and the consequent difficulties in deciding which resources to include or exclude, the figures quoted here are accurate to only 5 or 10%. This is however, quite sufficient for our purposes. It should be noted that in an earlier draft of this paper, a rather wider definition of 'running costs' was used, which is why the figures shown in this table are slightly smaller than those quoted previously.

[b] The 1978 exchange rates of 3.8 German Marks and 4.2 Dutch Guilders to the British Pound are used in these calculations.

[c] This includes the cost of the salaries for approximately 30 astronomers and 10 Ph.D. students from the universities of Leiden, Groningen, and Utrecht, who regularly use the NFRA research facilities, together with the salaries of 15 support staff.

[d] This includes the cost of the salaries for approximately 15 researchers and students (and of their support staff) at the universities of Bonn, Hamburg, Bochum, and Tübingen, who regularly use the MPI telescope.

[e] I.e. after excluding time spent teaching.

Bank appears to be about 20% larger than Cambridge, and the Dutch group over 50% larger. In terms of running costs, the differences in scale between the groups are apparently more pronounced, probably because of the higher salaries in continental Europe, with the Dutch group costing about three times as much and the German group over four times as much as each of the British groups.

Table 4
Details of major items of capital expenditure at each radio astronomy observatory up to 1978

Centre	Details	Completion date	Approximate capital cost
Cambridge	1-mile interferometer	1964	£0.5 M
	5 km interferometer	1972	£2.1 M
Jodrell Bank	Mark I 250-ft telescope	1957	£0.7 M
	Mark II 125 × 83-ft telescope	1964	£0.3 M
	Mark III 125 × 83-ft telescope	1967	£0.1 M
	Mark IA conversion	1971	£0.5 M
	Mark VA design costs [b]	1974	£0.7 M
	Multi-telescope radio-linked interferometer (MTRLI) – Knockin telescope	1977	£1.9 M
NFRA	1.5 km interferometer	1970	£3.0 M [a]
	Addition of 2 movable telescopes to 1.5 km interferometer	1975	£0.9 M
MPI	100 m telescope	1971	£4.0 M [a]

[a] In converting the Dutch and German costs into Sterling, we have used the yearly average exchange rates quoted in National Institute of Economic and Social Research [42].

[b] This telescope, which was intended to be a national, rather than a solely Jodrell Bank facility, was not in fact built. Details of the project and of the decision not to proceed with it can be found in a report by the Public Accounts Committee [48].

11. Publications

The first of the output indicators we shall consider is numbers of publications. These are shown in table 6, together with the aggregate and average for the ten years up to 1978 for each centre. [41] Only articles in properly refereed, scientific journals and in published conference proceedings (i.e. publications generally recognised by scientists as being contributions to science) are included. Articles providing popularised accounts of scientific research, unpublished work, preprints, [42] and

---

[41] These figures are based on the complete publication lists provided for us by each of the centres, and they should therefore be reasonably accurate. If there are errors, they are unlikely to be more than about 5%.

[42] In radio astronomy, nearly all preprints are eventually published; they are therefore excluded in order to avoid "double-counting".

Table 5
Number of support staff in 1978 [a]

	Cambridge	Jodrell Bank	NFRA	MPI
Total number of staff at the centre	90	104 [d]	160 [b]	240 [c]
Number of support staff (excluding astronomy researchers)	57	65	110	165
Number of support staff per researcher (including students)	1.7	1.7	2.2	2.2
Number of support staff per effective researcher (excluding students)	4.7	3.3	3.0	3.0
Number of support staff per effective researcher (including students)	3.0	1.9	2.3	2.3

[a] As in the previous tables, the figures may be slightly inaccurate because of the difficulties in estimating what to include and what to exclude in cases where resources are shared with other research groups. For example, the staff of the Mullard Radio Astronomy Observatory at Cambridge is 84, but there are in addition some employees, such as cleaners, telephone switchboard operators, and librarians who are shared with the Cavendish Laboratory, Cambridge University, bringing the total to about 90. The overall accuracy of the figures in the table should, however, be sufficient for our purposes.

[b] This includes 30 university astronomers and 10 students at Leiden, Groningen, and Utrecht, who carry out their research primarily on the NFRA facilities, together with an estimated 15 support staff (secretaries, cleaners, etc.) at these universities.

[c] This includes 15 astronomers and students at Bonn, Hamburg, Bochum, and Tübingen, together with an estimated 8 support staff at these universities.

[d] This excludes 16 staff formally employed to provide services to Jodrell Bank visitors (in the Concourse Building, restaurant, and Arboretum), a number of whom also provide some support services to the radio astronomers.

Table 6
Numbers of publications

	Cambridge	Jodrell Bank	NFRA	MPI
Publications in				
1946	1	1		
1947	1	5		
1948	6	11		
1949	2	3		
1950	5	8		
1951	6	13		
1952	9	14		
1953	3	13		
1954	7	20		
1955	13	18		
1956	10	19		
1957	18	20		
1958	10	7		
1959	6	24		
1960	18	24		
1961	21	20		
1962	19	20		
1963	17	22		
1964	12	20		
1965	22	17		
1966	30	11		
1967	43	15		
1968	48	22	15	
1969	40	21	16	16
1970	32	22	22	45
1971	32	21	41	41
1972	43	27	37	35
1973	33	23	65	57
1974	47	16	55	47
1975	40	21	66	65
1976	40	16	57	78
1977	46	18	92	87
1978	45	12	91	79
1969–78 aggregate	398	197	542	550
1969–78 average	40	20	54	55

laboratory reports are all excluded.

From these figures, one can see that the level of scientific production of the Cambridge group grew fairly steadily until the late 1970s when it levelled off. The slight "dip" in output in the early 1970s coincides with the period during which a large proportion of the group's effort was devoted to bringing the new 5-km interferometer into operation (see table 4). In the case of Jodrell Bank, the level of production grew until the end of the 1950s, during which time it was consistently higher than at Cambridge; after this, it dipped in the mid-1960s (when the Mark II and Mark III telescopes and multi-channel digital spectrometer were built), and has again declined more recently as a major effort has been put into the new multi-telescope radio-linked interferometer (see table 4). However, even after allowing for these temporary fluctuations, the average level of scientific production of Cambridge has still been considerably higher than that of the Jodrell Bank group over the ten years up to 1978.

As we have already seen, the numbers of publications is but a partial indicator of scientific pro-

BEST COPY AVAILABLE

gress. While the difference between the publication rates for Cambridge and Jodrell Bank *may* indicate a difference in scientific progress, it may also be accounted for by other factors. For example, it may be related to the different types of radio telescope and research interest at the two centres. Alternatively, there may be substantially different publication practices in the two groups with the result that Jodrell Bank publications on average make a more substantial impact in terms of their contribution to scientific knowledge. The latter possibility is examined in the next section on citations, but the former can be considered here by looking at the output of the Dutch and German groups.

Table 7
Numbers of publications relative to staffing and funding levels [a] in 1978 [b]

	Cambridge	Jodrell Bank [c]	NFRA	MPI
No. of papers per astronomy researcher (excluding research students)	2.6	0.5 (0.8)	2.3	1.3
No. of papers per astronomy researcher (including research students)	1.4	0.3 (0.5)	1.8	1.1
No. of papers per effective researcher [d] (excluding research students)	3.8	0.6 (1.0)	2.5	1.4
No. of papers per effective researcher (including research students)	1.6	0.4 (0.6)	1.9	1.1
Approximate cost per paper	£15 k	£75 k (45 k)	£30 k	£45 k

[a] The likely errors in these figures are probably somewhere between 5 and 10% (see note a to table 3 and note a to table 5.

[b] These are the figures for a single year only (1978), and it it is conceivable that they might be untypical of the decade 1969–78. However, for the reasons mentioned in note 38, it is unlikely that figures for the four centres would show a very different pattern in previous years, at least once the two continental centres began operating at full strength.

[c] The figures in parentheses are those obtained for Jodrell Bank if 1978's publication rate of 12 papers is assumed to be a temporary "dip", and the figure of 20 papers (the ten-year average) is substituted as being a more typical value.

[d] I.e. after excluding time spent on teaching (see table 2).

During the ten years up to 1978, the production of papers by both the Dutch and German groups has increased rapidly, finally reaching a level about twice that of the Cambridge group and more than four times that of Jodrell Bank. This high volume of production for the two continental groups can be partly explained in terms of their higher funding and staffing levels. If one allows for this (see table 7), it appears that the production of the Cambridge and the Dutch groups has been relative high – certainly higher than that of the German and Jodrell Bank groups (even after allowing for the latter's recent diversion of effort into instrument-building – see note c to table 7). Whether the smaller number of publications for the latter is compensated for by each publication representing, on average, a greater contribution to scientific knowledge is considered in the next section.

12. Citations

Figures on the numbers of citations to *all* previous research work by each centre have been calculated, using their complete publications lists and the *Science Citation Index*. These are detailed in table 8. The number of citations to Cambridge work has increased by over 150% during the period 1966–1978, with a slight pause to this growth evident in the 1974 figures. Table 9, giving figures for citations to work published over the previous four years, suggests that this temporary fall can probably be attributed to the diversion of effort at the beginning of the 1970s away from research and into the commissioning of the new 5-km interferometer.[41] Both the Dutch and German groups also show a rapid growth in the numbers of citations to their work, reaching, in 1978, a total similar in magnitude to that for Cambridge. In the

---

[41] The effects of this concentration on instrument-building show up even more clearly in table 11, which indicates that the citation rate per Cambridge publication dropped markedly in the early 1970s. During this period many of the staff were involved in building, testing, and commissioning their new telescope, and, while the publication rate of the group as a whole fell only slightly (largely kept up by some highly productive PhD students), the citation rate for each paper declined dramatically.

BEST COPY AVAILABLE

Table 8
Citations to all previous work [a]

	Cambridge	Jodrell Bank	NFRA	MPI
1966 citations	300	420	–	–
1970 citations	860	410		
1971 citations			210	
1972 citations				320
1974 citations	810	490	380	550
1978 citations	1380	470	1360	1030
1978 citations to 1969–78 work only	1120	340	1340	1030

[a] It is difficult to attach any estimate to the likely errors in these figures. If they were of the order of $\sqrt{n}$, this would correspond to percentage errors of between 3 and 7%.

Table 10
Rates of self-citation (SC) and "in-house" citation (IHC) in 1978

	Cambridge	Jodrell Bank	NFRA	MPI
Number of citations analysed in sample [a]	217	200	200	200
No. of SCs	9	4	8	9
No. of IHCs	32	28	36	30
Percentage of SCs	4%	2%	4%	5%
Percentage of IHCs	15%	14%	18%	15%
Total % of SCs and IHCs (likely error)	19% (± 3%)	16% (± 3%)	22% (± 3%)	20% (± 3%)

[a] This sample of 1978 citations was chosen on a random basis.

case of Jodrell Bank, however, the number of citations to previous work has scarcely changed during the 12 years studied. Similarly, table 9 shows that the most recent work by the Cambridge, Dutch, and German groups has all been far more highly cited than that by Jodrell Bank.

In the earlier theoretical discussion, it was argued that citation rates should be regarded as a partial indicator of scientific progress. This being so, we must now examine the extent to which the lower citation rate for Jodrell Bank reflects a smaller overall contribution to scientific progress than those of the other groups, or is merely a product of certain "other factors". But what might such "other factors" include? Can, for example, the low citation rate be attributed to weaker links with the international community of radio astronomers, less of whom therefore cite Jodrell Bank work? Scientists in both Holland and Germany commented, in interviews, that *both* the Cambridge and the Jodrell Bank groups tend to be somewhat "closed" in their attitudes towards outsiders, but they claimed that this tendency is more pronounced at Cambridge than at Jodrell Bank. Hence, it would seem difficult to attribute the latter's low citation rate to this cause. Can the differences in numbers of citations for each group be attributed to differing rates of self-citation and "in-house" citation [**] amongst the centres? A sample of about 200 1978 citations for each centre was examined, and table 10 shows that there is little significant difference between the four centres in the frequency with which they cite their own group's work.

Yet another possibility is that the total number of citations for the work of a group is very largely determined by its publication rate. We can examine whether this is a satisfactory explanation by looking at the average number of citations to each paper. Table 11 shows the figures for the citations to work published in the preceding four years, divided by the number of publications produced in that time, i.e. the average number of citations per publication (CPP). From table 11, it can be seen that the Cambridge group, while having a far lower overall output of papers than the two European groups, nevertheless achieves a significantly higher number of citations per publication. This suggests that the number of citations is not just a function of numbers of publications.

Table 9
Citations to work published in the previous four years. [a]

	Cambridge	Jodrell Bank	NFRA	MPI
1970 citations	340	300		
1971 citations			210	
1972 citations				320
1974 citations	330	220	480	340
1978 citations	550	190	780	610

[a] See note a to table 8.

[**] I.e. citations by a scientist to the work of colleagues in his or her group.

Table 11
Citations per publication (CPP) for work published in the previous four years [a]

	Cambridge	Jodrell Bank	NFRA	MPI
CPP in 1970	3.3	2.5		
CPP in 1971			2.2	
CPP in 1972				2.3
CPP in 1974	2.1	2.5	2.4	1.9
CPP in 1978	3.2	2.8	2.4	2.0

[a] The likely errors in these figures probably range between 3 and 7% (see note a to table 8).

There would, therefore, seem to be some justification for arguing that, for the radio astronomy centres in this study, citation rates do provide a useful partial indicator of scientific progress. In the case of total citation rates, the figures certainly *suggest* that the Cambridge, Dutch, and German groups have made rather more impact on the advance of knowledge than Jodrell Bank in the period 1969–1978, with Cambridge having done particularly well in relation to its size. The figures for citations per publication again suggest that, of the four centres, Cambridge papers have generally had the largest impact (with the exception of those at the start of the 1970s – see note 43); in this respect, Jodrell Bank seems to have done slightly better than the Dutch and German groups, although the *total* impact of all its papers is apparently still much smaller because of its lower publication rate. However, until we take account of the results of peer evaluation, we cannot determine with any confidence whether such differences in citation rates are significant or not, and whether figures on citations say anything about the overall contribution of each centre.

Table 12
Highly cited papers published between 1969 and 1978 [a]

	Cambridge	Jodrell Bank	NFRA	MPI
No. of papers cited 15 or more times in 1 year [b]	12 [c]	1	6 [d]	1 [c]
No. of times these highly cited papers received 15 or more citations in 1 year	23	1	13	1
No. of papers cited 20 or more times in 1 year	4 [c]	0	3 [d]	0 [c]
No. of papers cited 12 or more times in 1 year	19 [c]	3	7 [d]	2 [c]

[c] Excludes papers based on work carried out elsewhere by astronomers before moving to the centre.

[d] Includes one review paper, and one paper describing the new synthesis telescope.

[a] Rigorous efforts have been made to ensure that these figures are as accurate as possible. It is unlikely that any of the values in the first two rows is wrong by more than 1 or 2 units. Errors of this magnitude would have no effect on the conclusions drawn in the text concerning the distribution of highly cited papers between the centres.

[b] Out of the total of approximately 1700 publications produced by the four centres between 1969 and 1978, only 20 (i.e. about 1.2%) have been cited 15 or more times in any one year.

Before considering the peer-evaluation results, however, there is one further argument that could be levelled against the preceding citation analysis, and which needs to be scrutinised. It is argued by some (e.g. Cole and Cole [31]) that it is not the great mass of publications (each constituting but a minor increment to the sum of human knowledge) which contributes most to scientific progress, but a small number of key papers, each of which has a very great impact on the advance of knowledge. If this is the case, then the figures for total numbers of citations, and for citations per publication, may not reveal which centres have managed to produce such key papers and which have not. However, we can establish which centres have produced the most highly cited papers and, although (as was seen earlier) there are problems in using citations for comparing individual papers, this should give some indication as to the distribution of key papers between the four centres.[45] Figures for approximately the top 1% most cited papers (cited 15 or more times in any one year) are shown in table 12. It can be seen that between 1969 and 1978 the Cambridge group produced twelve such papers, and the Dutch group half that number, while the two other groups managed only one each. Some of these publications were highly cited only for a short time, after which their impact appears to have diminished rapidly; the second row of table 12 takes account of the variation between these short-lived papers and those that have been highly cited over many years by showing the number of times these papers managed to achieve 15 or more citations in a year. Again a similar pattern emerges, although the differences between the four groups are, if anything, further accentuated. (Lest it be imagined that this distribution of highly cited papers is in some way the result of choosing an arbitrary value of 15 citations a year as the dividing line between highly cited and less highly cited papers, table 12 also includes the numbers of papers cited 20 times or more, and 12 times or more, in each case yielding a similar distribution between the four centres.)

---

[45] One might suppose the number of highly cited papers to be at least a partial indicator of the number of "discoveries" made by a research group, and there is indeed some evidence to support this supposition from the peer-evaluation interviews described later. In these, astronomers were asked to identify the principal "discoveries" or significant contributions to scientific knowledge made by each of the research groups over the ten years up to 1978. There was a high degree of consensus among the interviewees on at least the major contributions, and these generally coincided with the most highly cited papers. Moreover, at one of the two centres with very few highly cited papers, many of the astronomers freely admitted that the centre had been responsible for no major discovery during the ten-year period.

13. Peer evaluation

In a series of approximately 70 interviews with virtually the entire scientific staffs of the two British groups, and with about one-third of the Dutch and German groups (including all the most senior astronomers), scientists were asked to identify the main contributions to scientific knowledge made over the period 1969–1978 by their own centre and by eight other major radio astronomy centres.[46] This group included all the radio observatories considered by the astronomers in our survey as being the foremost centers in this field. They were then requested to rank these centres in order according to the magnitude of their contributions to scientific knowledge made over the ten-year period.[47] We subsequently investigated the consistency of the evaluations produced within each centre, and the consistency between centres. In particular, we looked for systematic differences between groups, such as a tendency to over-rate one's own centre.

The average rankings (on a scale between 1 and 9[48] because there are nine centres) obtained at each centre are shown in table 13. A number of comments can be made about the results. First, there is a fairly high degree of consistency in the results obtained both *within* each centre, and *between* the centres. Indeed, if one averages the results for all four centres, the great majority (31 out of 36[49]) agree with the average rankings for all four centres to within one unit or less.

Secondly, there is *some* tendency for scientists to over-rate the scientific progress made by their own group, but while this is significant in two of the four centres,[50] it is not as marked as might have been expected, and does not appear to be sufficiently large to cast doubts about the overall significance of the peer-evaluation results.

Thirdly, it is possible to detect a slight influence of certain particularistic factors on these rankings. For example, Cambridge ratio astronomers tended to rank interferometer centres slightly more highly,

---

[46] Besides the four centres already discussed, these included the American radio telescopes at Arecibo, Caltech, and the National Radio Astronomy Observatory (NRAO), the French centre at Nançay, and the Australian big dish at Parkes. We have, as yet, not looked at publication and citation indicators for five of the nine centres, but hope to do so in the future to see how the results correlate with the results of our peer evaluation. This would then constitute a study of a large part of the world radio astronomy community, and the results would be of relevance not only to science policy, but also to the sociology of science.

[47] As might be expected, there was a certain reluctance to do this, generally on the grounds that, because each centre is involved in slightly different work, it is misleading to rank them on a single scale. Nevertheless, most respondents (88%) did agree to take part in this ranking exercise: some felt able to rank each centre in order from 1 to 9, while others preferred to group the centres into categories of "first class", "second class", "third class", etc.

[48] If two centres were ranked first-equal, they were given the average ranking of 1.5. Similarly, if three were ranked equal first, they were given the ranking of 2 (the average of 1, 2 and 3); and so on.

[49] Four of the remaining five differed by only 1.1 units, and the remaining one, where the discrepancy was 2.2, was a self-ranking – i.e. the ranking by one centre of its own position on the scale.

[50] In addition, the senior MPI staff tended to slightly over-rate themselves, but this was cancelled out by the more junior staff who, if anything, tended slightly to under-rate the scientific progress of the group. This difference between the evaluations of senior and junior staff at the German center may be related to a certain dissatisfaction amongst the latter about the organization and orientation of the group's research.

Table 13
Rankings of nine radio astronomy centres [b] obtained by peer evaluation (PE)

	PE at Cambridge (n = 11) [a]	PE at Jodrell Bank (n = 19) [a]	PE at NFRA (n = 13) [a]	PE at MPI (n = 18) [a]	Average rankings for all four centres	Average rankings excluding self-rankings
Arecibo	7.8	7.3	7.0	6.7	7.2 (±0.4)	7.2 (±0.4)
Caltech	4.6	6.0	6.7	5.3	5.7 (±0.8)	5.7 (±0.8)
Cambridge	2.0	3.1	1.9	1.6	2.2 (±0.6)	2.2 (±0.6)
Jodrell Bank	5.7	3.8	6.1	6.1	5.4 (±1.0)	6.0 (±0.2)
MPI	6.3	4.8	4.3	5.3	5.2 (±0.7)	5.1 (±0.8)
Nançay	8.7	8.9	8.9	8.9	8.8 (±0.1)	8.8 (±0.1)
NFRA	2.5	3.3	1.6	2.7	2.5 (±0.6)	2.8 (±0.3)
NRAO	3.2	4.3	2.8	2.6	3.2 (±0.7)	3.2 (±0.7)
Parkes	4.0	3.6	5.7	5.6	4.7 (±0.9)	4.7 (±0.9)

[a] n = sample size.

[b] Ranked in terms of the magnitude of their contributions to scientific knowledge between 1969 and 1978. The figures in parentheses in the final two columns indicate the root-mean-square variations between the rankings given by the different centres.

and some of the big-dish centres less highly, than average, suggesting that the type of equipment and type of research at each centre may affect the peer evaluations. Similarly, radio astronomers who had previously worked at one of the other centres tended to rank that centre more highly than average. For example, some of the American astronomers working in the Dutch and German groups ranked the American radio astronomy centres slightly more highly than average. These examples illustrate one of the potential problems with the use of peer evaluation as an indicator of scientific progress – that scientists may evaluate more highly those centres whose work is most familiar to them. However, our results suggest that, while this may be true for some individuals, the effects of these particularistic factors are barely significant when using fairly large samples of peer reviewers. [31]

The high degree of consistency in the peer-evaluation results, together with the relatively small effects of self-rating and other particularistic factors, gives some support for the belief that peer evaluation, at least when carried out with a large representative sample of scientists from carefully matched groups, can yield a reasonably good indication of the scientific progress made by research groups. Now, given that some confidence can be placed in the peer-evaluation results, what do the results actually show? The average rankings of the nine groups are shown in fig. 3, [32] these averages having been obtained after excluding self-rankings on the ground that these may sometimes produce a small but nevertheless significant effect on the results. From this figure it can be seen that the Cambridge and Dutch groups, together with the National Radio Astronomy Observatory (NRAO) in the United States, are clearly perceived as having been world-leaders in radio astronomy over the period 1969–1978. They are followed by a group of four other observatories, including Jodrell Bank and Bonn, with the former apparently in the lower half of this group. The two remaining groups were seen as having achieved significantly less

---

[31] The American astronomers working in Holland and Germany ranked two of the three American observatories slightly more highly than average, but the differences are scarcely significant. Arecibo was ranked 6.3 (±1.3) compared with the average figure of 7.2 (±0.4) and NRAO was ranked 2.8 (±1.3) compared with the average of 3.2 (±0.7). The third American centre at Caltech was ranked below average (at 6.4 (±1.8) compared with the average of 5.7 (±0.8)), but again the difference is probably not significant. Because of the relatively small size of this systematic deviation, it seems doubtful whether the overall peer-evaluation results would be substantially altered by including radio astronomers from countries other than those covered here. Nevertheless, it would be interesting to include the evaluations of French, American, and Australian radio astronomers, and this may be attempted in the future.

[32] It must be emphasised that this is *not* a graph; it is merely a pictorial representation of the numerical results in the right-hand column of table 13. The nine centres have been arranged in the order in which they were ranked solely to bring out the differences perceived by radio astronomers between the top three observatories, the second group of four, and the remaining two centres.

BEST COPY AVAILABLE

scientific progress over this period.[11] Of the four centres with which we are principally concerned, it can be seen that the Cambridge group is regarded as having been the most successful of the four over the ten years up to 1978, while the Jodrell Bank and Bonn groups have apparently achieved rather less scientific progress (or that at least is the perception of most radio astronomers).

It is important to note that fig. 3 presents a static picture; it does not show changes in the relative positions of the centres over time. Besides ranking the observatories in terms of their performance over the ten-year period, we did ask radio astronomers to identify any changes in position *during* that period. From their responses it is evident that, at the start of the ten years, the Cambridge group was clearly perceived as being *the* world leader in radio astronomy, while the big dishes at Parkes and Jodrell Bank had also proved highly successful over the preceding period. However, in subsequent years, there appears to have been a significant improvement in the relative positions of NRAO and of the two newcomers, NFRA and MPI (in line with the changes over time noted in the indicators discussed earlier). This improvement came at the expense of Parkes, Jodrell Bank, and Caltech (which suffered from the increasing concentration of American radio astronomy resources on NRAO). Cambridge's lead was challenged first by the Dutch group when their new interferometer began operating, and is now about to face a severe challenge from NRAO's new Very Large Array. And perhaps in the near future, the MTRLI (see table 4), which was under construction in the 5 years up to 1978, may restore Jodrell Bank to something closer to its original position in the early 1960s. Certainly the number of publications produced by Jodrell Bank has increased significantly since this work was carried out (to 20 in 1979), and the extent to which this represents an increased contribution to scientific knowledge relative to that made by other radio observatories would be an interesting subject for future study.

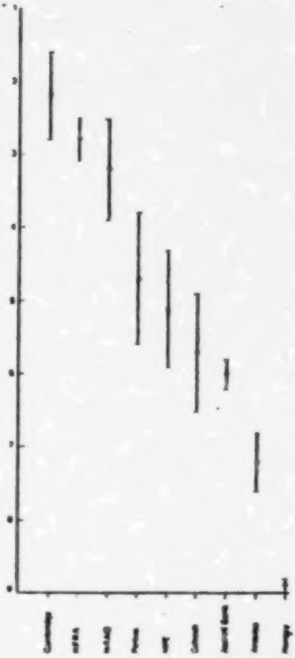


Fig. 3. Average rankings of nine radio astronomy groups for 1969–78 (excluding self-rankings). (The error-bars indicate the root-mean-square variations between the rankings given by the different centres.)

14. Discussion and conclusions

We began this paper by discussing the need for comprehensive, basic-science policies in modern industrial societies, where large amounts of state expenditure are now concentrated on a small number of "Big Science" centres. Perhaps the key

---

[11] In the course of the interviews, several other radio astronomy groups were mentioned as having achieved significant scientific progress. These included the groups at Bologna in Italy, Penticton in Canada, and Bell Telephone Laboratories in the United States. Some of these groups, if they had been included in our list of observatories, might have ranked as highly as the middle group of Parkes, Caltech, and Jodrell Bank, perhaps pushing some of these centres down the rankings. None, however, were seen as being in the same rank as the top three observatories.

AVAILABLE

problem in the past has been the lack of an adequate methodology for evaluating "success" in basic research. In the theoretical discussion in the earlier part of the paper, after analysing the limitations of such measures as publication counts, citation frequencies, "discoveries", and peer evaluation, we put forward a framework based on a set of partial indicators for comparing the scientific contributions made by research groups working in the same specialty. This was then applied to evaluate the performance of the two principal radio observatories in Britain – with NFRA and Bonn brought in as the major control groups. There are just two questions remaining to be addressed in this final section. First, to what extent are we justified in using the methodology described here to compare and to differentiate the scientific performance of basic research groups? And secondly, what policy implications can be drawn from the results obtained in this way?

The most important thing to note is that there is a high degree of consistency between the results obtained with four of the partial indicators. Numbers of highly cited papers and peer evaluation both suggest that the *total* contributions of the Cambridge and Dutch groups have been relatively large; and, if we allow for the differences in scale of research activity between the centres, the two partial indicators based on publications per researcher and citations per paper [54] suggest that the research productivity of the Cambridge group has been the highest of the four centres. This *convergence* in the results suggests [55] the selection of the control groups for comparison with the two British centres has been such that the effects of the "other factors" on the partial indicators have been kept relatively small, and hence that some reliability can be placed in assessments based on these four indicators. So what do these four indicators suggest for each of the radio astronomy centres? The results of this study suggest the following conclusions for each of the centres:

(1) The Cambridge radio astronomers are perceived by their peers as having achieved as much scientific progress over the decade 1969–78 as any other group in the world. Although there are larger and better funded groups, this disadvantage is offset partly by the Cambridge scientists achieving a high publication rate per researcher (especially if one allows for the time taken up with undergraduate teaching), and partly by producing papers which, on average, have more impact on the advance of scientific knowledge. They have also produced a comparatively large number of highly cited papers, [56] a result which is consistent with the judgement of their peers that they have made several major discoveries.

(2) The Dutch radio astronomers have achieved as much, or almost as much, as their Cambridge colleagues in terms of contributions to scientific knowledge. As at Cambridge, each researcher on average produces a large number of publications, so that, overall, the group produces more papers than Cambridge. However, the peer-evaluation results imply that the total Dutch contribution to scientific knowledge is no greater than for Cambridge, which in turn suggests that each Dutch publication does not constitute such a substantial contribution to knowledge, a conclusion apparently supported by the figures on citations per publication. Moreover, the group's research, although funded more generously, has resulted in significantly fewer, highly cited papers and, according to their peers, fewer discoveries than the Cambridge work.

(3) The large number of radio astronomers at Bonn ensures that, despite a somewhat lower publication rate per person, the group as a whole produces a large number of scientific

---

[54] The two other indicators – total numbers of publications and citations – do not yield results consistent with these four, and are therefore apparently not very reliable indicators of scientific progress. For example, both suggest that the German group has achieved as much as the Cambridge and Dutch groups. The reason is that numbers of publications reflect the level of scientific activity and production of a research centre rather than the scientific progress it is making, while a large total number of citations can be generated by publishing a large number of papers, even if each makes little impact on the advance of scientific knowledge.

[55] Convergence does not, of course, "prove" that the indicators are "reliable". There is still a possibility that, despite all the precautions taken, they are "unreliable" but linked together (see note 31), which is why conclusions can only be suggested rather than unreservedly drawn.

[56] 3.0% of Cambridge papers achieved 15 or more citations in a year, compared with 0.9% for the Dutch group, 0.5% for Jodrell Bank, and 0.2% for Bonn.

BEST COPY AVAILABLE

papers. The overall scientific progress made by the group is not, however, as great as for the Cambridge and Dutch groups, at least as judged by peer evaluation. This can be interpreted as implying that each German publication, an average, has a rather smaller impact on the advance of knowledge, a conclusion again in line with the figures for citations per publication. In addition, the Bonn group has not produced very many highly cited papers, nor many "discoveries" in the opinion of their peers.

(4) In the case of Jodrell Bank, the publication rate for the group as a whole is relatively low compared to that for the three other observatories, even after allowing for time taken up with undergraduate teaching and for the diversion of effort into the new interferometer project. This somewhat lower level of production is partly offset by producing papers of higher impact than those of the Dutch and Germans. Overall, however, the contributions to scientific knowledge by Cambridge and Dutch astronomers are judged by their peers to be significantly greater than those of their Jodrell Bank colleagues, and those of the German group may be slightly larger too, although, given that in terms of effective number of researchers (see table 2) the German group is over twice that of the Jodrell Bank team, any difference between the scientific progress made by these two groups may be mainly the result of this difference in scale of activity. There is certainly no difference evident in the numbers of highly cited papers – neither group has performed very well in this respect relative to NFRA and, in particular, to Cambridge.

It is not our task in this paper to make specific recommendations about individual centres. That is the job of science policy administrators responsible for the funding and organization of basic research. The decisions they reach will be dependent upon other factors such as the available funding, their analysis of the reasons why certain centres have achieved less than others, and their judgements as to whether it is possible to devise measures to improve the performance of the less successful centres. For example, one factor which has apparently influenced the scientific performance of radio observatories is the type of instrumentation employed. The indicators based on numbers of highly cited papers and on peer evaluation both suggest that the interferometer centres have been more successful than the big-dish centres over the period in question. What remains unclear is the extent to which this is a "real" effect (in the sense that interferometers have actually contributed more to scientific progress than big dishes), and the extent to which it is an artefact of the indicators employed. However, in view of the fact that the sample of radio astronomers used in the peer-evaluation process was not obviously biased towards interferometers (we actually interviewed rather more big-dish than interferometer astronomers), the peer-evaluation results do suggest that the effect is more real than artificial, and that the period between 1968 and 1978 should consequently be regarded as the era of the interferometer in radio astronomy.

Another factor that can influence the scientific performance of a research centre is the fraction of time its research staff devote to instrument-building, a fraction that may vary cyclically as new telescopes are brought into operation. To a certain extent, the relative performance of Jodrell Bank over the period 1969–78 may be linked to their concentration on instrument-building during much of this period.

However, we shall leave a discussion of the reasons for the differences in scientific performance of the various centres to a later paper, and return to the main objective of this paper, which has been to address two questions: (a) can basic research be assessed? (b) more specifically, can significant differences in the research performance of radio astronomy centres be identified? We would contend that the evidence presented in the paper is sufficient to justify a positive answer to both these questions. However, it should perhaps be emphasised that we have been concerned with the assessment of *past* performance, while policy-makers are clearly interested in the *future* performance of research groups. Nevertheless, past performance, although by no means the only factor, is one of the best indicators of future performance, particularly for Big Science where it is virtually impossible to set up a new centre or a major research programme overnight. We therefore see it as crucial that those who have to judge the respective claims for funds by different research groups, do so on the basis of full information on the recent performance of those groups. Although further work is obviously needed, we believe that this paper goes some way in demonstrating how that information can be obtained. It establishes a methodology that, while not able to compare *directly* research centres in different specialties, is able to identify those centres that are amongst the international leaders in their own specialties, and which, if other factors are equal or indeterminate, should be given a relatively high priority in their claims for research funds.

References

[1] F.M. Andrews (ed.) *Scientific Productivity: The Effectiveness of Research Groups in Six Countries* Cambridge University Press, Cambridge, 1979).

[2] A.E. Bayer and J. Folger, Some correlates of a citation measure of productivity in science, *Sociology of Education* 39 (1966) 381–90.

[3] J.L. Chan, Organisational consensus regarding the relative importance of research output indicators, *Accounting Review* 53 (1978) 309–23.

[4] H. Chang and D. Dieks, The Dutch output of publications in physics, *Research Policy* 5 (1976) 380–96.

[5] D.E. Chubin and S.D. Moitra, Content analysis of references: Adjunct or alternative to citation counting, *Social Studies of Science* 5 (1975) 423–41.

[6] K.E. Clark, *America's Psychologists: A Survey of a Growing Profession* (American Psychological Association, Washington, DC, 1957).

[7] S. Cole and J.R. Cole, Scientific output and recognition: A study in the operation of the reward system in science, *American Sociological Review* 32 (1967) 377–90.

[8] J.R. Cole and S. Cole, The Ortega hypothesis, *Science* 178 (1972) 368–75.

[9] J.R. Cole and S. Cole, *Social Stratification in Science* (University of Chicago Press, Chicago, 1973).

[10] J.R. Cole and S. Cole, Citation analysis, *Science* 183 (1974) 32–33.

[11] D. Crane, Scientists at major and minor universities: A study of productivity and recognition, *American Sociological Review* 30 (1965) 699–714.

[12] D.L. Croom, Dangers in the use of the Science Citation Index, *Nature* 227 (1970) 1173.

[13] D. Dieks and H. Chang, Differences in impact of scientific publications: Some indices derived from a citation analysis, *Social Studies of Science* 6 (1976) 247–67.

[14] C. Freeman, Measurement of output of research and experimental development, *UNESCO Statistical Reports and Studies no. 16*, ST/S/16, Com 69/XVI-16A (1969).

[15] E. Garfield, Citation indexes in sociological and historical research, *American Documentation* 14, no. 4 (1963) 29–31.

[16] E. Garfield, Citation indexing for studying science, *Nature* 227 (1970) 669–71.

[17] E. Garfield, Citation analysis as a tool in journal evaluation, *Science* 178 (1972) 471–79.

[18] E. Garfield, What scientific journals can tell us about scientific journals, *IEEE Transactions on Professional Communication* PC-16, no. 4 (1973) 200–02.

[19] E. Garfield, Citation and distinction, *Nature* 242 (1973) 485.

[20] G.N. Gilbert and S. Woolgar, The quantitative study of science: An examination of the literature, *Science Studies* 4 (1974) 279–94.

[21] H. Inhaber and K. Przednowek, Quality of research and the Nobel Prizes, *Social Studies of Science* 6 (1976) 33–50.

[22] J. Irvine and B.R. Martin, The economic effects of Big Science: The case of radio astronomy, *Proceedings of the International Colloquium on the Economic Effects of Space and Other Advanced Technologies, Strasbourg, 28 - 30 April 1980* (Ref. ESA SP-151, Paris), 103–16 (1980).

[23] N.C. Janke, Abuses of citation indexing, *Science* 156 (1967) 892.

[24] N. Kaplan, The norms of citation behaviour: Prolegomena to the footnote, *American Documentation* 16 (1965) 179–84.

[25] T.S. Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press, Chicago, 1970).

[26] J. Larabi, Mesure de l'efficacité des laboratoires de recherche fondamentale sélectionnes par le Centre National d'Etudes Spatiales, *Revue Francaise d'Informatique de Recherche Operationnelle* 3 (1969) 103–12.

[27] J. Larabi, Note sur l'efficacité des laboratoires de recherche fondamentale sélectionnes par le CNES, *Le Progres Scientifique* 137 (1970) 4–18.

[28] S.M. Lawani, Citation analysis and the quality of scientific productivity, *Bioscience* 27, no. 1 (1977) 26–31.

[29] D. Lindsey, Production and citation measures in the sociology of science: The problem of multiple authorship, *Social Studies of Science* 10 (1980) 145–62.

[30] J. Maddox, Is the literature worth keeping?, *Bulletin of the Atomic Scientists* 19, no. 9 (1963) 14–16.

[31] J. Margolis, Citation indexing and evaluation of scientific papers, *Science* 155 (1967) 1213–19.

[32] B.R. Martin, Cognitive and social locations: Their role in the processes of discovery and evaluation within science (Department of Liberal Studies in Science, Manchester University), mimeo (1977).

[33] A.J. Matheson, Centres of chemical excellence?, *Chemistry in Britain* 8 (1972) 207–10.

[34] A.J. Meadows and J.G. O'Connor, Bibliographical statistics as a guide to growth points in science, *Science Studies* 1 (1971) 95–99.

[35] J.D. McGervey, Citation analysis, *Science* 183 (1974) 28–31.

[36] I.I. Mitroff, and D.E. Chubin, Peer review at the NSF: A dialectical policy analysis, *Social Studies of Science* 9 (1979) 199–232.

[37] M.J. Moravcsik, Measures of scientific growth, *Research Policy* 2 (1973) 266–75.

[38] M.J. Moravcsik, A progress report on the quantification of science, *Journal of Scientific and Industrial Research* (India) 36 (1977) 195.

[39] M.J. Moravcsik and P. Murugesan, Some results on the function and quality of citations, *Social Studies of Science* 5 (1975) 86–92.

[40] M.J. Moravcsik and P. Murugesan, Citation patterns in scientific revolutions, *Scientometrics* 1 (1979) 161.

[illegible] AVAILABLE

[41] C.R. Myers, Journal citations and scientific eminence in contemporary psychology, *American Psychologist* 25 (1970) 1041–48.

[42] National Institute of Economic and Social Research, *National Institute Economic Review* 86 (1978) table 25.

[43] Is your lab well cited? *Nature* 227 (1970) 219.

[44] More games with numbers, *Nature* 228 (1970) 698–99.

[45] University of Houston expels professor, *Nature* 279 (1979) 278.

[46] A.L. Porter, Citation analysis: Queries and caveats, *Social Studies of Science* 7 (1977) 257–67.

[47] Price, D. de S., *Little Science, Big Science* (Columbia University Press, New York, 1963).

[48] Public Accounts Committee, Expenditure on design for proposed Mark VA radio telescope, *Fifth Report from the Committee of Public Accounts*, Session 1975–76, HC 556 (HMSO, 1976 London), pp. xxxi–iv, and 260–70.

[49] E. Shils (ed.), *Criteria for Scientific Development: Public Policy and National Goals* (MIT Press, Cambridge, Mass., 1968).

[50] R. Smith and F.E. Fiedler, The measurement of scholarly work: A critical review of the literature, *Educational Record* (1971) 225–32.

[51] D. Sullivan, D.H. White and E.J. Barboni, The state of a science: Indicators in the specialty of weak interactions, *Social Studies of Science* 7 (1977) 167–200.

[52] N. Wade, Citation analysis: A new tool for science administrators, *Science* 188 (1975) 429–32.

[53] A.M. Weinberg, Criteria for scientific choice, *Minerva* 1 (1963) 159–71.

[54] J.H. Westbrook, Identifying significant research, *Science* 132 (1960) 1229–34.

[55] R.D. Whitley, Communication nets in science: Status and citation patterns in animal physiology, *Sociological Review* 17 (1969) 219–33.

5 What Direction for Basic Scientific Research?

John Irvine and Ben R. Martin*
Science Policy Research Unit, University of Sussex

INTRODUCTION

The future for basic scientific research[1] currently seems far from secure. In most OECD countries, rapid growth rates in public expenditure have long since given way to static budgets and even cuts. National science budgets have not been exempt from this trend. Initially this gave rise to demands for more effective science policy - for restructuring and 'rationalization' of the organization of scientific research. When budgetary cuts were later made, these tended with certain notable exceptions (for example, social science research) to be imposed 'across the board'; science policy-makers had no acceptable way of systematically evaluating which research activities were more (or less) in need of support than others. In the absence of any explicit data concerning the nature and magnitude of the wide range of outputs from different scientific activities, such policies inevitably receive criticism as arbitrary and divorced from any coherent, long-term strategy for basic research.

New mechanisms for decision-making are needed if the direction of basic scientific research is to be put on a satisfactory basis in the future. But what are the various types of decisions currently facing those concerned with science policy-making? What methods and criteria can be used in making them? What problems are associated with these methods? And to what extent can such problems be overcome, so as to yield more acceptable and hopefully more effective policies for scientific research in the future? This chapter sets out to suggest ways of answering such questions.

---

* No order of seniority implied (rotating first authorship). The authors are Fellows of the Science Policy Research Unit, University of Sussex, where they work on a range of issues concerning government policies for scientific research. The work reported here was funded by the Social Science Research Council. Thanks are due to various colleagues, in particular Jay Gershuny, Dot Griffiths, John Krige, Ian Miles, Geoff Oldham, Keith Pavitt, Tim Shallice, and Tom Whiston for their comments on earlier drafts of this paper.

RESOURCE ALLOCATION IN BASIC SCIENTIFIC RESEARCH

In allocating resources to basic research, decisions are required at a number of levels. Starting with a total state r&d budget, these involve asking:

1. How much overall should the state spend on basic science compared with other areas of r&d such as the more applied sciences and engineering?
2. How should this overall budget be distributed between different (basic) sciences such as physics, chemistry, biology, and the social sciences?
3. What should be the distribution between different specialties within each science (e.g. between optical and radio astronomy, high-energy physics, solid-state physics, etc., within the physical sciences)?
4. Within each area of science, how should funds be allocated between different types of activity (postgraduate-training, university research, centralized research in national laboratories, international collaboration)?
5. How much should be allocated to each research centre, group, or individual, working within a given specialty, both to optimize current research activities and ensure the maintenance of an effective research effort in the future?

PROBLEMS IN RESOURCE ALLOCATION

The coming crisis in basic research was to some extent anticipated in the early 1960s by the eminent physicist, Alvin Weinberg. Weinberg pointed out the seeming absence of any underlying rationale to resource-allocation patterns in basic science. In what became a seminal paper in the science policy literature, he distinguished between two sets of criteria - internal and external[2] - that he advocated should be more widely applied to determine the relative levels of support for different scientific activities. Internal criteria relate to the impact the research is likely to have on the advance of scientific knowledge within the field con-

BEST COPY AVAILABLE

cerned. External criteria, in contrast, refer to the extent to which the research is seen as leading to wider benefits to other scientific fields and to society, for example, in the form of new ideas and techniques for industrial application. While there are problems in this distinction - 'internal' and 'external' criteria are not mutually exclusive, and their definition is to some extent a political exercise - it does nevertheless provide a useful analytic framework with which to approach a discussion of the problems of resource allocation.

At a very general level, it is clear that external criteria do play some role in determining the overall size of the science budget, and (to a lesser extent) the distribution of resources between broad fields of scientific activity (e.g. to the natural sciences, social sciences, or biological sciences). Such external criteria, however, have largely been implicit in the decision-making process[3], and, as we shall see below, have tended to reflect the values and commitments of certain established interest groups. In contrast, the distribution of resources within fields and specialties appears, in the Western world at least, to have been made largely on the basis of internal criteria. Such decisions have tended to be the prerogative of the scientific community; the assumption has been that only scientists themselves are competent to judge which research is likely to contribute most to scientific knowledge, and hence to decide between the competing claims of different researchers for funds. In the words of Polanyi:

> 'So long as each allocation follows the guidance of scientific opinion, by giving preference to the most promising scientists and subjects, the distribution of grants will automatically yield the maximum advantage for the advancement of science as a whole'.[4]

The fact that such a justification entirely begs the question of what is meant by 'promising', and how it is judged, has all too frequently been ignored.

The principal mechanism for the organization of basic scientific research has thus been based largely on internal regulation. Politicians and civil servants have normally been content to leave such matters to the scientific community, provided, of course, that the resultant activities are in general accord with the external criteria implicit in decision-making at government level. However, the pressures of living within a level budget have highlighted a number of problems with internal mechanisms for the regulation of scientific activity. These problems have been accentuated in many countries by government attempts to introduce stricter and more explicit external criteria at increasingly lower levels in the decision-making process. It is therefore time that stock is taken of the strengths and weaknesses of internal and external regulation of science as currently constituted. Only on this basis can we consider how these mechanisms for resource allocation might be improved.

REGULATION BY INTERNAL CRITERIA

Regulation of scientific research by internal criteria has traditionally been based almost entirely on peer-evaluation. Peer-evaluation has been described as:

> 'a kind of science advice involving select members of the community, all of whom act - to a greater or lesser degree - as gatekeepers. These gatekeepers help to regulate the flows of information and fiscal resources through the community by directing, impeding, and expediting flows based upon judgements of quality and merit, allegiances and biases and, probably, on sheer caprice as well'.[5]

Despite criticism from those scientists unable to pass the 'gatekeepers' for reasons which may well be other than lack of competence (for example, for personal, political, bureaucratic or institutional reasons) the opinion of most scientists[6] seems to be that peer-review has worked reasonably well in the past as a regulatory mechanism for resource allocation, although there have been occasional doubts about its merits, for example in the 1960s at the time of the debate on Weinberg's criteria for scientific choice (see note 2). In part, this is because flexibility to accommodate demands for resources from most established scientific groups and disciplines was built into the system by the rapid expansion of national science budgets between 1945 and the early 1970s. It is clear that there is far less strain on a system administering steadily increasing budgets than one oriented towards restructuring and reducing expenditure, where 'gatekeepers' are put under severe pressure to abandon their professed disinterestedness and openly defend (or attack) individual, institutional and specialty interests.

Indeed, it could be argued that the whole infrastructure for allocating research resources on the basis of peer-review is now under such pressure that it is in danger of breaking down. There are three main reasons for this. First, the long period of post-war growth in basic research activities has led to the entrenchment of particular interests in decision-making structures. To a large extent, this problem has its historical origins in the immediate post-war period during which there was an unprecedented expansion of basic research activity (for example, nearly 500 per cent growth in Britain during the five years between 1945 and 1950). At that time, politicians attached great strategic importance to science (because of the central role they perceived scientists to have played in the Second World War), and were therefore willing to set aside the funds required. Many of the scientists who worked on high-technology warfare techniques such as radar and nuclear bomb construction subsequently moved into developing associated areas of natural science such as radio astronomy and high-energy physics.[7] Their prominence in the war effort, and the social and institutional links they had developed with politicians and civil servants, made it relatively easy to obtain funds to build large research facilities such as radio telescopes and accelerators. These research areas were therefore able to grow rapidly, and to become what is now known as 'Big Science'. Today, for example, one major high-energy physics centre (CERN, the European laboratory in Geneva) has an annual budget of over 300 million dollars; and one new accelerator (the LEP electron-positron collider) alone is currently planned to cost approximately 500 million dollars.

Most importantly, as the level of funds for Big Science grew (to over 50 per cent of the basic research budget in some countries during the 1960s), so did the representation of their respective research communities on the bodies responsible for distributing the science budget. In the case of Britain, the situation is now such that, within the Science and Engineering Research Council, the organization responsible for funding research in the natural sciences, one subsection of physics, nuclear physics, has its own Board, as does astronomy and space research, while all other areas of science - physics, chemistry, biology, mathematics and computing - have only one Board between them. The result is the situation depicted in Table 1: Big Science consumes almost double the resources of all the remaining natural sciences together, whilst constituting only a small proportion of total scientific activity, at least in terms of numbers of doctoral studentships and postdoctoral research fellowships.

With resource distribution so heavily dependent on decisions made by scientists themselves, there has inevitably been a tendency for certain early established sets of priorities and research interests to become 'frozen into' the decision-making structure. To those on the periphery of this structure, the system can all too easily appear to be little more than a means of ensuring that those funded generously in the past continue to be supported generously in the future.[8] Whether or not this judgement is accurate, it does point to a key problem associated with the scientific community itself having main responsibility for decisions on the distribution of the basic science budget. This is that the strength of the case for additional funds made by a particular specialty or science depends to a great extent on the strength of its representation on decision-making bodies. (It also depends on the associated access to the media and political channels - avenues of influence the more 'glamorous' Big Sciences have not been slow to exploit.) Thus, it is possible that certain older established areas of research, which in all likelihood stand little chance of producing substantial contributions to scientific knowledge, will continue to be heavily supported because they are well represented in the committees determining funding priorities. In addition, such areas have often been able to institutionalize large fractions of their research support into longer-term commitments that can be reduced only very slowly, if at all. Conversely, scientists in a new and potentially important field may sometimes find difficulty (unless established scientists lend their support) in eliciting funds because there is no relevant committee (the committees tend to reflect prevailing patterns of scientific activity) or because they are poorly represented.

The second major problem concerns the effects in certain specialties, most notably the Big Sciences, of concentrating research activity within fewer and fewer institutions. As these areas have developed, so they have come to demand increasingly sophisticated and expensive research facilities (astronomers now have access to space-borne telescopes, for

BEST COPY AVAILABLE

Table 1: Some Features of British Science and Engineering Research Council Funding of Basic and Applied Science, 1981/2

Area of Expenditure[1]	Budget	Studentships and Fellowships current in 1981/2	
	£M	Doctoral Research Studentships	Post-doctoral Fellowships
BIG SCIENCE:			
Nuclear Physics	44.0	176	9
Astronomy, Space and Radio	37.1	223	24
Total Big Science	81.1	399	33
ALL OTHER BASIC AND APPLIED SCIENTIFIC ACTIVITY (coming under the Science Board)	44.6	3790	111
of which:			
SERC establishments[2]	16.7	-	-
International contributions	7.3	-	-
Other domestic facilities	1.0	-	-
Research grants to:			
Biological sciences	6.8	1499	54
Chemistry	7.0	1439	32
Mathematics	0.7	351	13
Neutron Beam Physics	0.2	9	0
All other physics	4.5	468	12
Other sciences	0.4	24	0
Central support schemes, Projects and administration[3]	16.8	227[4]	6[4]
TOTAL SERC ACTIVITY IN BASIC AND APPLIED SCIENCE[1]	142.5	4416	150

Notes:

1. This excludes SERC expenditure on postgraduate awards and

advanced training grants, and all engineering activity.
2. Much of this is for the support of two SERC establishments recently re-oriented from high-energy physics, but still working in areas requiring the provision of large, expensive research facilities (e.g. synchrotron radiation and neutron physics) whose non-inclusion as 'Big Science' is therefore a matter of definition.
3. This excludes an allowance of £3.3 million for engineering activity.
4. These are funded by the Joint SERC/SSRC Committee, the Polytechnics Committee, NERC, etc.

Source: Adapted from data in *Report of the Science and Engineering Research Council for the year 1981-82* (Swindon: SERC).

(astronomers have now access to space-borne telescopes, for example). Since it is no longer possible to provide separate research facilities for all groups working in a field, there has been a trend towards siting major instruments and equipment in a small number of central laboratories - either regional or national centres, or, in the case of the most expensive areas of research, international centres. In Britain, for example, seven large research centres accounted in 1981/2 for approximately £106 million (see Table 2), i.e. some 65 per cent of all SERC expenditure on basic and applied scientific research. This is over three and a half times the total allocated in the form of peer-reviewed research grants to academic scientists.[9] While these centres are used by large numbers of university scientists, most academic research is not dependent on such central facilities. And since it is university departments which are responsible for training the new generations of science students for government and industry, there is, not surprisingly, some discontent[10] with this policy of concentrating resources on a few centres and their associated user-groups. Moreover, the problem is not confined to physical science: in medical science, for example, there are similar disputes over the relative emphasis that should be given to research institutes and units compared with university departments.

It is, however, with the political effects of such concentration that we are most concerned here. In a specialty like high-energy physics, where formerly there were a number of university groups, each operating their own

BEST COPY AVAILABLE

WHAT DIRECTION FOR BASIC SCIENTIFIC RESEARCH

Table 2: Concentration of British Science and Engineering Research Council Expenditure on Basic and Applied Scientific Research on Large Central Facilities, 1981/2

	£M
1. CERN (High energy physics)	22.5
2. Rutherford Appleton Laboratory (Mixed facilities, including central laser facility, spallation neutron source, high-energy physics support, and space)	40.0
3. Institut Laue-Langevin (Neutron beam facilities)	7.2
4. Daresbury Laboratory (Synchrotron radiation source, and nuclear structure facility)	16.2
5. European Space Agency	8.3
6. Royal Greenwich Observatory (Optical and infra-red astronomy)	8.5
7. Royal Observatory Edinburgh (Optical and infra-red astronomy)	3.4
Total SERC expenditure on these 7 facilities	106.1
Total SERC expenditure on all science (excluding engineering but including postgraduate awards)	164.2

Note: 1. The figure for the Rutherford Appleton Laboratory excludes about £8.7 million which was spent on engineering research.

Source: *Report of the Science and Engineering Research Council for the Year 1981/2* (Swindon: SERC).

BEST COPY AVAILABLE

accelerator, there are now at most two or three main research facilities available to the scientific community of any one country. This poses a threat to the traditional means of determining research priorities; the successful operation of peer-evaluation depends upon the existence of a constituency of 'disinterested' peers capable of providing relatively independent expert judgements. In other words, there must exist sufficient scientists familiar with the research area where support is being considered, but whose own material circumstances will be unaffected by the outcome of the decision-making process. When the number of distinct research groups working in a specialty is relatively large, this condition is at least approximately met. However, in the Big Sciences, when the question of allocating resources to a major centre is being considered, nearly all peers will be a user either of that centre (and so will benefit from a positive decision), or of another rival centre (whose own chances of obtaining further funds may increase with a negative decision on the first centre). Alternatively, the users of the two or three centres may reach a tacit agreement to support each others' grant and capital-equipment applications (regardless of what they might privately think of the merits of those applications), taking it in turns to seek funds. They thereby present a facade of unity and consensus on priorities within the specialty, which is always a distinct advantage when competing for funds with other specialties. Instead of the 'free market' of scientific ideas - all competing with an equal chance of funding - the notion on which scientists have traditionally based their view of peer-evaluation as a neutral disinterested process[11], there has been a tendency towards a situation of 'oligopoly', in which a few large scientific centres and interest groups are able to exert a dominant influence on funding-mechanisms. The open and 'objective' scientific community (which, like the 'free market' of classical economics, is probably more a historical myth than an accurate representation of reality) has given way to a decidedly hierarchical system where the distribution of resources increasingly seems to obey the maxim, 'to them that hath shall be given'.

A third and directly related problem concerns the general ineffectiveness of peer-evaluation as a mechanism for restructuring scientific activity. This particular problem has only really become apparent since the science budgets of

BEST COPY AVAILABLE

Western countries ceased to grow. Science is inherently dynamic in nature, with new areas of activity continually emerging and other older areas reaching a point where further research is subject to diminishing returns. The long-term vitality of the basic research system, and indeed its capacity to provide the knowledge and skills to help solve the pressing economic and social problems facing the world today, is dependent on the ability to support promising new areas of science. Consequently, it is crucial to maintain flexibility in the system even in a period of level funding. When science budgets were growing substantially each year, it was not necessary to curtail existing research activities; instead annual budget increases could be channelled to fund promising new activities, leaving researchers in 'declining' areas to retire or move into other areas of science of their own volition. However, with a level budget, this option is no longer available; if support for promising new areas is to be found, and found promptly, then reductions must generally first be made elsewhere. These might involve decreasing the support for declining areas, or for research groups which, although working in flourishing research areas, have not succeeded in making substantive advances commensurate with the level of resources they consume.

The problem is that, while peer-evaluation may be a relatively effective mechanism for deciding between promising new areas of research, it is far less satisfactory when it comes to identifying declining areas and groups. This may be partly through lack of experience, since it is not a task scientists have often had to perform in the past. However, there are other deeper-rooted institutional and social factors severely limiting the use of peer-evaluation in this way. The institutional limitations arise because, although there is obviously greatest scope for savings in the case of heavily-funded research activities, these are precisely the areas where reduced expenditure is likely to be most strenuously resisted by senior scientists strategically placed on science-policy committees.[12] The social structure of science also limits the use of peer-evaluation for identifying research groups or centres which have proved relatively unsuccessful in their previous research. A scientist asked to judge whether the funds of a research group in her or his speciality should be reduced, is likely to know some of the members of that group personally, perhaps having collaborated with them in previous research, met them regularly at conferences, or interacted with them at the relevant learned society. For that scientist, a decision as to whether or not to recommend a cut in funds is exceedingly difficult to make, jeopardizing as it does the future livelihood of colleagues.[13]

It is certainly far harder than (and qualitatively different from) deciding whether to support giving a particular research group additional funds, the main type of decision that had to be made in more affluent times, and one where a negative outcome merely meant that new equipment could not be purchased or additional researchers could not be recruited. Consequently, when circumstances have forced peer-reviewers into choosing among candidates for reduced support, the criteria used for selection may often have had more to do with social or political factors than past or likely future scientific performance.[14] Furthermore, it should be noted that peer-reviewers tend to be recruited most heavily from the higher status institutions; the cuts they recommend may therefore result in a strengthening of traditional status hierarchies, when perhaps what is most needed is a redistribution of resources benefiting some of those research groups previously less well-endowed.[15]

To sum up our discussion of the three main problems facing peer-review: first, in a period of level funding, where resources for existing research activities must often be reduced if new areas are to be supported, peer-evaluation seems to be proving increasingly ineffective as a mechanism for establishing priorities, and establishing them sufficiently expeditiously. Because representation on scientific decision-making bodies reflects previous patterns of resource distribution, there is a tendency towards the reproduction of that distribution among both specialties and institutions. Secondly, because research in many areas of science demands ever more sophisticated centres in each specialty, it is becoming increasingly difficult to obtain the disinterested judgements on which peer-review depends. Finally, peer-evaluation, at least in its traditional form, may be inappropriate for identifying declining research areas or groups. New mechanisms may be required if scientific research is to be more effectively and acceptably regulated.

BEST COPY AVAILABLE

WHAT DIRECTION FOR BASIC SCIENTIFIC RESEARCH

REGULATION BY EXTERNAL CRITERIA

As economic growth rates in the industrialized countries have declined, so their governments have come to inspect more carefully all areas of public expenditure. In particular, this has led to increasing demands that basic research should produce more direct and practical benefits to society, and not just 'good science'. Similar demands have also been made by scientific pressure groups at the radical end of the political spectrum such as the British Society for Social Responsibility in Science, with their quest for a 'Science for the People'.

To some extent, this can be seen as a reaction to the ethos prevailing within the basic research community under which science is seen as the neutral and disinterested pursuit of knowledge for its own sake. No one would deny that basic science does result in a variety of important practical benefits (and some disbenefits), including skilled researchers entering the labour market, technological 'spin-off' to equipment suppliers, new ideas and techniques of longer-term utility to industry, as well as cultural contributions in the form of greater awareness of the workings of the social and material world. However, most scientists see these simply as by-products of basic research - highly important but unplannable. And, in view of this, they argue that there is no reason why the principal criteria used to determine the distribution of research funds should not be internal ones, decided upon and operationalized by the scientific community. This ethos, moreover, serves to protect the power of scientists to determine their own activities (within certain limits set by the social and institutional structure of science), enabling them to avoid what are seen as the negative consequences of state-planning of basic research. To support this view, scientists point to examples such as the Lysenko affair to illustrate the horrors associated with state-planning of research. This is often coupled with an expression of the belief (and it is no more than that - there is little or no relevant evidence to support it) that, if too much importance is attached to external criteria, basic research will gradually be pushed further and further towards the applied end of the pure-applied spectrum.

One might have expected that, with increasing governmental pressure to adopt better defined external criteria in funding decisions, the scientific community would adopt a slightly more self-critical attitude towards the question of where basic research stands today - in particular concerning the current distribution of research efforts between different sciences and specialties, and the assumptions underlying that pattern of distribution. However, where scientists have attempted to do this - for example, in COSPUP, the Committee on Science and Public Policy set up in the United States in the early 1960s,[16] they have met with little success. One possible explanation is that researchers in each scientific field have been more concerned with protecting the interests of that field than arriving at agreement on the relative priorities to be given to different fields in deciding between their competing claims for public funds. Instead, scientists have concentrated on producing reports that focus on the claims of individual fields.[17] In the United States, there has been a succession of reports - the Ramsey Report on High Energy Physics, the Whitford Report on Astronomy, the Westheimer Report on Chemistry, the Pake Report on Physics, and so on. Each spelt out at great length the immense scientific opportunities confronting these fields, eloquently putting forward reasons why the nation must avail itself of these unique opportunities, even if this did entail some additional expenditure on the field. However, apart from Weinberg's discussion of possible criteria for scientific choice, very little attention has been given to the question of how one might attempt to establish priorities between fields. In particular, there have been few attempts to evaluate and compare systematically, even at a general level, the various outputs associated with different areas of scientific activity. Yet in the absence of any such evaluation, the application of external criteria to determine research policy will inevitably run the risk of yielding decisions that can be too easily dismissed (perhaps rightly so) as biassed or politically motivated.

The argument of scientists that basic research should not be subject to regulation by external criteria, however, ignores the fact that certain criteria, often implicit, have shaped and continue to shape scientific decision-making. These implicit criteria all too frequently go unrecognized and unexamined by the scientific community. Examples include the notion that certain areas of research are more 'fundamental' than others, or that they are more 'strategically important'. These are often given as justifications for the heavy support of particular research activities[18], yet rarely is any serious consideration given to precisely what is meant by such terms as 'fundamental' or 'strategically important'. However, the implicit nature of such criteria does not mean they can be ignored; any decision on the distribution of resources, to research or any other activity, implies the existence of criteria, whether explicitly taken into account or not. For example, the figures in Table 3 on the distribution of state funds for r&d show how markedly the pattern of resource allocation can vary across even relatively similar European countries. In the case of Britain, there are various political and institutional factors which account for the extremely high level of support for military r&d. Similarly (and notwithstanding the claims of some scientists to the contrary) external criteria must play a role in determining the relative size of the basic-research budget, justifying a level of support well over three times that for r&d oriented to industrial growth (even though in international terms British basic research appears comparatively underfunded).

In the same way, external criteria can also be seen as implicit in the distribution of the basic research budget to the different areas and specialties, even if no formal attempt is made to apply such criteria in the decision-making process. Table 4 contains figures on the distribution of funds in Britain to the five research councils, the bodies responsible for supporting most pure and applied research, as well as engineering. It is interesting here to compare the figure of £44 million for expenditure on nuclear physics (from Table 1) in 1981/2 (most of which went on high-energy physics) with that on the different research councils. High-energy physics, which after all is only one sub-field out of many in the discipline of physics, received more support than the whole of agricultural science and nearly as much as all areas of environmental science. On the face of it, it would seem somewhat difficult to explain such a distribution pattern in terms of the operation of internal criteria alone - are advances in our knowledge of high-energy physics to be valued so much more highly than those in agricultural or environmental science, and, if so, on what grounds?[19] It seems far more plausible to suggest

Table 3: State Priorities in R&D Expenditure of Five European Countries by Socio-Economic Objectives, 1980

Priorities	Belgium	France	Germany	Italy	United Kingdom
Percentage of Total Government Funding:					
Agriculture	4.8	3.9	1.9	4.1	4.1
Industrial Growth	14.9	9.2	9.9	17.4	6.2
Production of Energy	8.7	7.8	13.9	22.9	6.7
Transport, Telecommunications	1.4	2.9	-[2]	0.4	0.7
Urban and Rural Planning	2.0	1.4	3.8	0.8	1.0
Environment Protection	2.9	1.1	-[2]	1.0	0.9
Health	13.0	4.5	6.2	4.5	1.6
Social Development Services	11.1	1.3	4.0	2.2	1.1
Earth and Atmosphere	3.7	3.0	2.8	2.1	0.9
Advancement of Knowledge (i.e. Basic Research)	31.8	22.7	42.8	35.7	20.5[3]
Civil Space	5.6	4.5	4.4	6.2	2.1
Defence	0.3	37.2	10.2	2.7	54.2
Not Specified	0.0	0.4	0.0	0.0	0.9
Total (%)	100.0	100.0	100.0	100.0	100.0
Total in Million US$ (Current Prices)	721	7147	9324	1809	5680
Total as Percentage of GDP	0.61	1.10	1.19	0.47	1.08

Notes: 1. Figures are for 1980/1 (2) Figures included elsewhere (3) Excludes general university funds in the Social Sciences and Humanities.

Source: *OECD Science Resources Newsletter*, No. 6, Summer 1981, Table 2.

Table 4: Distribution of British Government Support to the Research Councils, 1981/2

	£M
Agricultural Research Council (ARC)	42.1 (9.7%)
Medical Research Council (MRC)	101.7 (23.3%)
Natural Environment Research Council (NERC)	54.3 (12.6%)
Science and Engineering Research Council (SERC)	216.8 (44.8%)
Social Science Research Council (SSRC)	20.7 (4.8%)
TOTAL	435.6 (100%)

Source: *Appropriation Accounts 1981-2*, Vol.7, Class X, pp.2-3 (London: HMSO).

that certain external criteria helped give rise to this distribution apparently so skewed in favour of Big Science.

In the case of high-energy physics, it is clear that one such criterion was the assumption of a link with nuclear energy. In the 1940s and 1950s, it was assumed by politicians and the public at least, that the discoveries made with high-energy physics accelerators would in some way help solve the problems involved in producing nuclear energy. The scientists concerned made little effort to disabuse the public of this assumed link. Indeed, the titles they chose for the organizations responsible for high-energy physics research (the European Organization for Nuclear Research, CERN; the National Institute for Research in Nuclear Science) can only have reinforced this supposed

BEST COPY AVAILABLE

link. No attempt was made to publicize the fact that, right from the early 1950s, high-energy physics and nuclear-energy research bagan to pursue distinct and very separate paths of development. In fact, in the United States, high-energy physics is still funded by the Department of Energy, and in France by the Atomic Energy Commission, even though it has long seemed clear that the direct links of this specialty with energy research are virtually non-existent. It was this assumption that high-energy physics would provide background knowledge for future nuclear-energy production which gave it an apparent strategic importance over other basic research specialties in the 1950s and 1960s, and which was used to justify it receiving a large fraction of the total basic-research budget in Western countries.[20] Later on, political criteria came to the fore. When the costs of research in this field escalated to such a level that most nations could no longer afford to build and maintain national facilities, high-energy physicists could, for example, argue to their governments that continued membership of CERN, even if expensive, provided very visible proof that they were 'good Europeans', an argument that was, for example, particularly appealing to many British politicians in the late 1960s.[21]

The problem now is the absence of appropriate mechanisms for restructuring the basic research budget - for redistributing resources (when this is deemed necessary) away from certain specialties to others. As we noted earlier, this is not a task easily performed using internal criteria because of the ossification of the peer-review system. Hence, the current distribution of the basic research budget to the different areas or specialties in science is probably more a reflection of priorities that were fixed early in the post-war period than the result of conscious planning on the basis of explicit criteria in subsequent decades.

Certainly, there have been some attempts in recent years to give greater attention to explicit external criteria. However, to the outside observer, these seem to have been more concerned with justifying the existing distribution of resources than seeking to establish future priorities on a systematic, explicit, and, as far as possible, impartial basis. The pressures towards greater public accountability have naturally been greatest in the most expensive areas of scientific activity, in particular space research and high-energy physics. As a response to this, NASA, ESA (the European Space Agency) and CERN have all promoted studies concerning the economic spin-off derived from their work. The resulting reports[22] have claimed that the overall economic benefits have been considerably greater than the investments made in these high-cost areas of research. These studies may have been good 'public relations', helping to secure the funds needed to continue such research, but a number of reservations must be noted in connection with this approach to evaluating external benefits. First, these studies have in large part been methodologically dubious. Using data provided by firms benefiting from equipment contracts, they yield figures for rates of return on investment that at first sight appear very precise and highly impressive (270 per cent in the case of ESA). However, closer inspection of the methods used to derive these numbers makes it far from clear what significance can be attached to them - for example, there is no mention of the opportunity costs associated with such investment. Secondly it is probably only the largest research groups which can afford to commission such work, leaving other less capital-intensive areas of research at a severe disadvantage. Indeed, the finding that investment in ESA has led to economic returns over a period of 5 to 10 years 2.7 times greater than the costs, even if the methodology used to derive it were less suspect, is of limited relevance unless one also has to hand figures on the rates of return for other research activities. Moreover, one would also need to introduce into the calculation the possible negative effects on established rival firms which are not awarded these contracts. For such firms, the consequence of a competitor being given a lucrative procurement contract may be a declining market share, perhaps even threatening their long-term financial viability. Thirdly, and perhaps more importantly, there is a danger that goo great an emphasis on the immediate technological and economic spin-off from these forms of research activity will result in a neglect of activities relating to longer-term scientific problems, the solutions to which may yield benefits to society only in the distant future. Similarly those areas able to demonstrate immediate spin-off, by virtue, for example, of their need for high-technology equipment, may well attract attention (and funds) away from those activities where the benefits are more diffuse, although no less real.

BEST COPY AVAILABLE

There is also a fourth, more general issue raised by economic spin-off studies of the sort carried out by ESA. This concerns their insensitivity to the political and ideological problems inherent in any attempt to formulate and apply external criteria. Criteria such as 'economic benefit' are not neutral categories: 'economic benefit' accrues unequally, and by no means everyone would agree as to its value, particularly those who have to accept the associated disbenefits. The benefits are often concentrated on a relatively small number of firms, yet spin-off calculations rarely consider the question of who bears the cost (ultimately the public) and who benefits. It cannot necessarily be assumed that what is good for a few firms leads to general social benefit. In most studies attempting to apply external criteria to evaluate scientific activity, it is only the more readily quantifiable outputs that are focussed upon (as in traditional cost-benefit analysis). These are invariably economic in nature rather than social or political. Yet there are many factors relating to the impact of different forms of scientific activity other than just economic growth and profit, and these are systematically ignored in most studies. This is, of course, not just because they are more difficult to quantify. A more fundamental reason is that alternative sets of criteria may well embody different values that are less acceptable to politicians and the media. To take one example, in a recent study examining the economic impact of radio astronomy,[23] we found substantial industrial-training 'benefits' associated with this activity. From the beginning, there has been a considerable flow of trained graduates into high-technology industry, transferring with them the skills and techniques (in computing, image-analysis, etc.) developed during their apprenticeship in radio astronomy. Given that most previous studies in the area have focussed mainly on technological spin-off, this constituted, within certain limits, an interesting piece of policy research. These limits concern the values and ideology presupposed in such a notion as 'economic benefit': the data presented in this report could just as easily be interpreted as showing that the flow into industry was not just into any form of high-technology industry but principally into sectors associated with national security work - satellite construction, computer surveillance, remote sensing, and automated finger-print recognition, for example. It is thus possible to impose a set of alternative and largely conflicting values when applying external criteria to resource allocation under which contributions to military and police activities would be valued as a 'disbenefit' rather than a 'benefit'. Instead of concluding that radio astronomy had important social benefits because of its utility to industry, the study could instead be taken as providing further evidence of what are seen as the insidious links between much basic research and the military-industrial complex.

This example is only intended to highlight a general problem associated with studies which purport to present an 'objective' evaluation of external benefits associated with scientific research. This is that they ignore the fact that in modern industrial societies there exists a plurality of values. These reflect differing sets of political, cultural and economic interests, and are often in conflict. All too frequently, such assessments fail to recognize the value-laden nature of the assumptions underlying them. Instead they uncritically adopt one particular set of values which gives precedence to economic considerations over criteria relating to freedom, amenity, environment, health, and societal problems. Inevitably these studies tend to arrive at conclusions which justify and reinforce existing patterns of resource distribution within scientific research, rather than questioning whether a radically different set of priorities is perhaps required.

In short, the current pattern of r&d expenditure, even if it is not explicitly based on external criteria, does presuppose the existence of certain implicit criteria. By refusing to accept that external criteria have any legitimate role to play in the regulation of research activity, the scientific community avoids crucial questions being put concerning the future direction and use of science, leaving the present structure of values embodied in the distribution of funds to go unexamined and unchallenged.

REDIRECTING BASIC RESEARCH

It is clear from the above discussion that the decision-making system in science is currently in a state of crisis. Not only has the internal system for resource allocation in basic research grown increasingly rigid, but there exist no adequate means of linking scientific endeavour with socially determined priorities. Overcoming these problems will be

BEST COPY AVAILABLE

one of the main tasks facing science in the 1980s.

Perhaps the major obstacle to change is the political and ideological resistance of the scientific community, channelled through powerful organizational structures. The view put forward by Polanyi that 'the soil of academic science must be exterritorial in order to secure its control by scientific opinion'[24] is still the dominant ethos among scientists. Nevertheless, the need to restructure and redirect scientific activity is now being discussed by many national science councils, government ministries and international organizations. The NATO Science Committee, which has assumed the role of trying to promote change, argued in a recent 'Message to the Political, Industrial and Scientific Authorities of the NATO Member Countries' that

> 'the scientific community must unavoidably make sacrifices in agreeing to break down barriers, abolish established privileges, redistribute facilities, make international rather than national options, set aside the less promising and re-group all its activities within the framework of a more selective international community.'[25]

However, such a course implies a considerable degree of faith in the self-regulation of the scientific community, and probably over-estimates the ability of scientists to put their own house in order. Instead, some form of outside involvement may well be necessary to structure and catalyse change, particularly with regard to questions of resource allocation. The crucial issue here is the extent to which it is possible to integrate such initiatives with traditional mechanisms (especially peer-evaluation) in a way that the scientific community as a whole does not see as threatening to their own activities. For it is clear that outside participation in scientific affairs will be most effective if accepted by the scientific community in a spirit of enlightened self-interest.

The question of scientific as opposed to social control of science is, of course, far from new[26], and it is one that has generated a great deal of antagonism and misunderstanding in the past - both within and outside the scientific community. On the one hand, scientists have asserted, generally with a minimum or justification, that only they have the competence to decide on scientific affairs; on the other, there have been those who have attempted to treat science in the same way as technology, seeking to 'plan' scientific research so as to achieve broad social and economic goals in a way that denies the existence of any internal logic to scientific progress.

Our view is that it is now possible to go beyond such apparently irreconcilable positions, and to develop more effective means of regulating science from the point of view of facilitating scientific progress, while at the same time ensuring that overall scientific activity is kept broadly in tune with likely future societal demands for basic knowledge and techniques. To do so will mean 'opening up' science, providing the information to enable scientists from other disciplines, scientific administrators, politicians and the general public to formulate judgements on the inputs and outputs for each area of research. In short, science will have to become more openly accountable. This will inevitably involve reforming and extending the present systems for decision-making in science.[27] This, we believe, can be achieved in ways that need not in principle undermine the basic premises underlying scientists' traditionally voiced desire to exercise effective control over the planning and execution of research activity. There is space here only to outline in broad detail the main elements of a programme of change that would permit the redirection and restructuring of science in the 1980s.

First of all, it is clear that significant improvements are needed in the data available on the inputs to the different areas of scientific activity. At present, few countries have a clear idea of how much is spent, on what, and how this compares with the spending patterns of other countries (for example, data on the distribution of funds between disciplines and institutes, or between equipment, salaries, training, research grants, etc.) The United States has taken an important first step in this respect with the biannual publication by the NSF of *Science Indicators*, but there have been few matching initiatives elsewhere, despite the encouragement of the OECD Science and Technology Indicators Unit, which publishes the general statistical digest, *Science Resources Newsletter*. In the absence of more reliable data, particularly disaggregated data, it is extremely difficult to evaluate trands, to set priorities, to relate the inputs for different areas of science to the outputs, and hence to arrive at judgements of relative productivity. This will imply a much larger statistical effort on the part of most nations to monitor the distribution of research resources, but one that is warranted by the level of resources currently consumed by scientific research.

Secondly, it will be necessary to overhaul the system by which resources are allocated to different scientists, institutions and scientific specialties. Currently this is based almost entirely on peer-evaluation which, as we have seen, is facing various problems. Certainly scientists, by virtue of their expertise and specialized knowledge, are in principle best placed to identify the most timely and promising research, and the researchers able to execute the work most successfully. For this reason, peer-evaluation must remain at the centre of decision-making structures in basic science. Nevertheless, major improvements are necessary in peer-evaluation procedures, and in many cases these can be introduced quite easily. In particular, much wider use can be made of foreign peers to overcome the bureaucratic and organizational inertia of national interest groups, particularly those centred around large capital facilities. (In Britain, for example, foreign peers have not generally been used to help judge research applications in areas of Big Science such as high-energy physics and astronomy.) It is also possible to envisage alternative, more sophisticated methods of peer-evaluation, involving iterative processes with wider panels (including scientific 'lay-people' as well as peers for the area of research concerned) evaluating a range of proposals at any one time. Similarly, it may well be of value to develop more open evaluation systems, where the comments of referees are themselves made available for open scrutiny and criticism. The possibility of referees being challenged to defend their reasons for rejecting or accepting proposals may help alleviate the rather alarming situation documented by Cole *et al*,[28] who concluded from their recent study of peer-review in the United States that,

> 'the fate of a particular grant application is roughly half determined by the characteristics of the proposal and the principal investigator, and about half by apparently random elements which may be characterized as the 'luck of the reviewer draw''.[29]

These various problems with conventional peer-review mechanisms have been recognized in the Netherlands in particular, where some of the changes outlined here have already been adopted by the Organization for the Advancement of Pure Research (ZWO). All physics proposals are now judged by a jury consisting of physicists from a range of specialties, together with a number of chemists, mathematicians, and astronomers. The jury is an ad hoc one, being reformed anew each year. It does not meet, thereby avoiding the danger that its members may be unduly swayed by the arguments of a persuasive advocate for a particular proposal rather than by its intrinsic merits; indeed its members are not allowed to discuss proposals with each other. In the first round of assessment, jury members merely ask questions and make criticisms of proposed projects. At the same time, expert advice on each proposal is obtained (by mail) from four to six referees. The referees, at least two of whom are foreign, are not asked to recommend whether the proposal should be accepted or rejected, but they are asked to give reasons why they think the proposal is worthwhile or not. The referees' comments, suitably edited by ZWO in order to preserve anonymity, are shown to the grant applications, who have the right of reply to these criticisms and to those made by the jury. On the basis of all this evidence, the jury then grades the proposals in terms of four criteria - the ability of the researchers, the objectives of the work, the methodology, and an overall quality grade. Although the final decision as to which proposals should be funded rests with the relevant ZWO board, this two-stage Delphi technique has been found to yield a high degree of consensus amongst jury members, with the result that the jury's advice is nearly always followed.[30] Moreover, despite initial reservations, the Dutch scientific community has adopted the new procedure with some enthusiasm. Indeed, university engineers have since begun to press for a similar procedure to be adopted for applied science.[31]

In judging between research proposals competing for limited resources, one of the few indications of the success with which the proposed projects are likely to be carried out is the past performance of the research groups or institutions concerned. It is here that there is perhaps greatest scope for improvement to the decision-making process. Various techniques now exist for evaluating past research performance in a rigorous and systematic manner. These draw upon the methods commonly used by social scientists (for example, structured interviews to produce coherent sets of peer-evaluations, and attitude-surveys), and in particular the indicators developed in the 'science of science' - for example, publication counts and the frequency with which research publications are cited by other scientists.

Such evaluation methods are capable of yielding information of considerable use to decision-makers in that they supplement and complement the results from traditional peer-review procedures. The methods have been applied to a number of scientific specialties,[32] and have proved successful in identifying research centres which, despite being relatively generously supported, have failed to make scientific contributions commensurate with the funding received. Under conventional peer-evaluation procedures, knowledge of the relatively poor performance of such centres may be confined to a limited number of scientists or administrators, whose interests are such that they prefer not to disclose these facts and initiate wider debate in the scientific community. Consequently, those responsible for the distribution of scientific resources may occasionally have to make decisions on the basis of less than complete information.

At a time of static science budgets, regular systematic evaluations of the past performance of those large research centres which account for a high proportion of the basic science expenditure are essential. Only by identifying in a reliable, reproducible manner the less successful research activities is it possible to free the resources needed to support promising new research groups or specialties, and young scientists in particular.

Thirdly, much greater attention needs to be paid to the sociological, institutional and organizational problems associated with modern scientific research. While there exists a comparatively large literature on the factors contributing to successful technological innovation, there is a dearth of systematic information on the conditions for success in scientific research. Certainly, experience suggests that the 'ageing' of researchers, their restricted mobility, and blocked career-ladders are growing problems in many Western countries. Yet much more needs to be known about the effects of such factors as leadership and managerial style, the size and organization of research groups, and instrumentational obsolescence, on scientific performance.

In particular, very little is known about the process by which research groups emerge, evolve and mature, and the factors explaining the trend towards the concentration of resources on a few large centres. Is this trend inevitable and irreversible? If so, what are the policy implications, and what steps need to be taken to accommodate them? For example, can research councils like Britain's SERC adequately fulfil two roles at the same time - operating their own research centres, and acting as a funding agency for the entire nation's research effort - or are the two roles potentially incompatible? And when research council centres complete the task for which they were originally set up, how can they be closed or how can their efforts be redirected into new research fields given that they generally lack the flexibility associated with university research while at the same time they represent powerful interest groups acting upon research councils? It is clear that we need to develop alternative mechanisms for flexibly reallocating funds, personnel and equipment within the research system. We must learn how to restructure science in an effective yet humane way. Only by taking such action will it be possible to reorient activity to meet new needs - scientific and social - as and when they arise.

This brings us to the final point, the urgent requirement to develop mechanisms for ensuring closer integration of science with future economic and social needs. As was noted earlier, little is known about the precise nature of the linkages between different basic sciences and technology,[33] particularly those that can to some extent be predicted (for example, students entering the workforce with skills relevant to new technologies, such as superconductivity or microelectronics). This is clearly an area ripe for research. Methods need to be established for systematically identifying and comparing the likely social, economic and technological benefits (or costs) associated with different areas of research. While this sort of work is problematic, and though there is likely to be resistance to the application of such external criteria from sections of the scientific community, it must be re-emphasized that all we are advocating here is the replacement of the implicit external criteria currently responsible for shaping the distribution of the science budget by a set of explicit criteria arrived at as a result of public debate in which a wider range of interested parties have the opportunity and the information required to participate.

Again, institutional changes will be needed in the scientific decision-making system if external criteria are to be properly applied. In particular, more widely based committees representing a plurality of opinions, interests and values will be required. The 'information' provided to help make scientific decisions will always be subject to a number of different interpretations, and different groups with different values will identify and defend these, each interest group attaching its own political 'weights' to the various possible outputs from science. A healthy, well-run and open scientific system should have nothing to fear from an extension of the franchise in this way. A system dominated by certain organizational interests and characterized by a high level of secrecy surely has. We need to start formulating science-policy mechanisms that recognize the need for open debate, that admit the inevitability of healthy disagreement in a pluralistic system where values and objectives often conflict, and that force these values and objectives out into the open. Implicit assumptions underlying a particular science policy (for example, assumptions about the relative importance attached to internal criteria compared with external ones) could then no longer be hidden from the public gaze.

Secrecy, all too often the last refuge of the vested interest, would be replaced by greater openness from which, we would argue, a more acceptable and ultimately more effective distribution of scientific resources would result.

NOTES

1. We use the term 'basic scientific research' to refer to research carried out explicitly with the primary aim of creating new scientific knowledge, although this will often lead to subsequent social, economic and technological benefits.
2. A. Weinberg, 'Criteria for scientific choice', *Minerva* 1 (1963), pp. 159-71. The distinction we use later in this article is not identical with Weinberg's, since one of his external criteria ('scientific merit') is incor-

porated in our internal category.

3. It must be emphasized that we are talking here about basic research. Obviously, external criteria normally play a central role in decisions on the distribution of state funds to applied research - for example, to military R&D.
4. M. Polanyi, 'The republic of science: its political and economic theory', *Minerva* 1 (1962), p.61.
5. I. Mitroff and D. Chubin, 'Peer review at the NSF: a dialectical policy analysis, *Social Studies of Science* 9 (1979), p.200.
6. This statement and many of the other observations below are based on the results of interviews by the authors with over 400 scientists. However, because the interviewees were all working in Big Science, their views may not be wholly representative of scientists in general. For this reason, figures on the percentage of scientists holding one view or another have not been quoted.
7. Evidence for this comes from our interviews with radio astronomers and high-energy physicists (see note 6).
8. See note 6.
9. In 1981/2, the value of SERC grants to astronomy, space and radio, nuclear physics, and other science (but excluding engineering) was £31.8 million - see *Report of the Science and Engineering Research Council for the Year 1981/82* (Swindon: SERC), p.11.
10. See note 6.
11. Cf. M. Polanyi, *op.cit.*
12. This reaction has been aptly termed 'the closed-ranks defence reflex' in a recent NATO Science Committee pamphlet, *Research Management and Zero Growth* (NATO Scientific Affairs Division, Paris, 1981).
13. See note 6.
14. The strengthening of existing patterns of funding during a period of financial cutbacks is well illustrated in recent decisions made (in the summer of 1981) by the British University Grants Committee in their recommendations on how to spread substantial cuts across the university system. While the secrecy under which the UGC operates makes it unclear what performance indicators (if any) were used, the net effect was to concentrate the cuts mainly on certain newly established universities (most of whom did not have representatives or previous students on this Committee).

15. It should be stressed that we are not advocating that restructuring take place over-night. Scientists and their support staff cannot be treated simply as disposable resources, particularly in the current economic climate. Much more thought clearly needs to be given to the question of policies for restructuring scientific activity, including, for example, provision for retraining and greater institutional mobility.
16. This was a committee of the US National Academy of Sciences set up and chaired by George Kistiakowsky, a former presidential science adviser. It had a wide-ranging brief to study and pronounce upon any problem coming under the heading of science and public policy. For further details, see D.S. Greenberg, *The Politics of American Science* (Penguin Books, 1969, pp. 207-10).
17. One of the earliest such reports explicitly denied the possibility of making choices *between* scientific fields, and therefore arrived at the inevitable conclusion that each field should be given as much money as it needed: 'It is not possible to assign relative priorities to various fields of basic science nor should they be placed in competition. Each science, at any given time, faces a set of critical problems that require solutions for continued growth. Sometimes these can be acquired at little cost; sometimes larger expenditures of funds are needed. Hence, the cost may not reflect the relative value but rather the need. Each area must be funded according to these needs ....' (Piore Panel Report, *High Energy Physics Program*, 1958, p.138, emphasis added - quoted in Greenberg, *op.cit.*, pp.286-7).
18. See, for example, H. Bethe, 'High energy physics' (p.9 in L.C.L. Yuan (ed) *Nature of Matter: Purposes of High Energy Physics*, Brookhaven National Laboratory, BNL-888, 1965): 'High energy physics is undoubtedly today the frontier of physics. The discoveries in this field of study contribute most to the advance of our fundamental understanding of nature ... Solid state theory is still a very fruitful field ... [but] one could hardly claim that it advances our *fundamental* understanding of nature.' (Original emphasis).
19. Some might argue that high-energy physics, although expensive, is so important that it must be included in the 'portfolio' of research activities supported by highly industrialized nations. Even if this argument is accepted, it does not help in determining the precise

level of support the subject should receive. If all nations substantially reduced their support for high-energy physics, this might mean exploration of the quark structure of matter taking 20 years rather than 10. But it is far from obvious why such discoveries have to be made in 10 years rather than 15, 20 or 25.

20. Commenting upon the setting up of the National Institute for Research in Nuclear Science in 1957, an editorial in *Nature* (Vol.179, 27 April 1957, p.835) referred to high-energy nuclear physics as 'that highly competitive and economically vital field' (emphasis added), and went on to observe that 'the urgency of the demands on Britain's depleted reserves of energy places the field of nuclear science in a unique position of public concern'.
21. See the Parliamentary discussion reported in *Hansard HC* 768, columns 808-30, especially the speech by P. Kirk.
22. M.K. Evans, *The Economic Impact of NASA R&D Funding* (NASA CR 144351, 1976); J-P. Fitoussi et al., *Economic Benefits of ESA Contracts* (ESA/C(78) 104, 1978); H. Schmied, 'A study of economic utility from CERN contracts', *IEEE Transactions on Engineering Management* EM24, pp. 125-38. A non-quantitative but perhaps ultimately more convincing exposition of the likely benefits from one particular basic science (high-energy physics) is to be found in L.M. Lederman, 'Viewpoint from fundamental science', paper presented at the NSF Workshop on Basic Research and Development for the Advancement of Energy Science and Technology (mimeo, Fermilab, Batavia, Illinois, March 1982).
23. J. Irvine and B.R. Martin, 'The economic effects of Big Science: the case of radio astronomy', *Proceedings of the International Colloquium on Economic Effects of Space and Other Advanced Technologies, Strasbourg, 28-30 April 1980* (Ref. ESA SP-151, September 1980).
24. M. Polanyi (1962), *op. cit.*, p.67.
25. NATO Science Committee (1981), *op. cit.*, p.1.
26. See, for example, J.D. Bernal and M. Polanyi, 'Ought science to be planned? Two opposing views', *Bulletin of the Atomic Scientists* 5 (1949) pp.17-20.
27. Whether such reforms are possible in the absence of wider political and social changes is a debateable question, but one which we cannot address here.
28. S. Cole, J.R. Cole and G.A. Simon, 'Chance and consensus in peer-review', *Science* 214 (20 November 1981), p.885.
29. This finding was aptly summed up in the headline to the

New Scientist article about this work, 'God does play dice - with scientists' grants' (*New Scientist*, 19 November 1981, p.483).

30. See C. Le Pair, 'Decision-making on grant applications in a small country', *Scientia Yugoslavica* 6 (1980), pp. 137-43, for further details.
31. If such a procedure were to be adopted elsewhere and to become more widely accepted in the research community, then it would be but a fairly small step to broaden still further the membership of the jury to include non-scientists as well.
32. See B.R. Martin and J. Irvine, 'Assessing basic research: some partial indicators of scientific progress in radio astronomy', *Research Policy* 12 (1983), and 'Internal criteria for scientific choice: an evaluation of the research in high-energy physics using electron accelerators', *Minerva* 19 (1981), pp.408-32; and J. Irvine and B.R. Martin, 'Assessing basic research: the case of the Isaac Newton Telescope', *Social Studies of Science* 13 (1983), pp.49-86.
33. Some general discussion of the relationship between scientific research and innovation can be found in M. Gibbons and R.D. Johnston, *The Interaction of Science and Technology*, final report of a study carried out for the Economic Benefits Working Group of the Council for Scientific Policy (Department of Liberal Studies in Science, Manchester University, October 1972).

CERN: Past performance and future prospects

I. CERN's position in world high-energy physics

Ben R. MARTIN and John IRVINE *

Science Policy Research Unit, University of Sussex, Brighton BN1 9RF, UK

Final version received February 1984

In a series of three papers, we attempt to evaluate the past scientific performance of the three main particle accelerators at the Geneva-based European Organization for Nuclear Research (CERN) over the period since 1960, and to assess the future prospects for CERN and its users over the next ten to fifteen years.

We concern ourselves in this paper (paper I) with the position of the CERN accelerators in world high-energy physics relative to those at other large laboratories working in the field. We deal primarily with the period from 1969 to 1978, and attempt to establish how the experimental output from the three principal CERN accelerators, *taken as a whole*, compares with that from other major facilities. In undertaking this comparative evaluation, we draw on the method of "converging partial indicators" used in previous studies of three Big Science specialities.

In contrast, the second paper (paper II – Irvine and Martin [12]) focuses in detail on the scientific performance of each of the CERN accelerators *taken individually*. In particular, it asks, first, how the outputs from the CERN 28 GeV (giga or billion electron-volts) Proton Synchrotron compare with those from a very similar 33 GeV American accelerator at Brookhaven National Laboratory over the past two decades. Second, how great have been the experimental achievements of the CERN Intersecting Storage Rings in world terms? And, third, how do the outputs from the CERN 400 GeV Super Proton Synchrotron and from a rival US machine at Fermi National Accelerator Laboratory compare? Attempts are then made to identify the main factors responsible for determining the relative scientific performance of each CERN machine.

These factors are of relevance to the subject of a third paper (paper III – Martin and Irvine [20]) which sets out to assess the future prospects for CERN and in particular for LEP, the large electron–positron collider scheduled for completion in the latter half of the 1980s. What are the construction requirements (financial and technical) associated with LEP, and how easily will they be met? How does the scientific potential of LEP compare with that of other major accelerators under construction around the world? And, in the light of the previous record of the CERN accelerators, to what extent is this potential likely to be realized? The paper concludes with a discussion of the extent to which predictive techniques can be utilized in the formulation of scientific priorities, and of the problems in current science policy-making that such techniques might help address.

1. Introduction

> The Organization shall provide for collaboration among European States in nuclear research of a pure scientific and fundamental character, and in research essentially related thereto.

On September 29, 1954, the Convention for the Establishment of a European Organization for Nuclear Research came into operation. This had been signed the year before by the representatives of twelve states.[1] On the same date, the Conseil

---

* No order of seniority implied (rotating first authorship). The authors are Fellows of the Science Policy Research Unit, where they work on a range of issues connected with policies for basic and applied research. They gratefully acknowledge the support of the British Social Science Research Council in carrying out this research, and the help so freely given by large numbers of high-energy physicists. In particular, we must thank the senior management at CERN for allowing us full and open access to the laboratory during our research work. Thanks are also due to a number of colleagues at SPRU, and to the late Sir John Adams, Sir Clifford Butler, Professor John Dowell, Dr. Yves Goldschmidt-Clermont, Professor Tom Kibble, Dr. Leon Lederman, Dr. Owen Lock, and Dr. Paul Musset, for providing useful critical comment on earlier drafts of this paper. However, the conclusions remain the responsibility of the authors alone.

[1] The founding States were Belgium, Denmark, France, the Federal Republic of Germany, Greece, Italy, the Netherlands, Norway, Sweden, Switzerland, the United Kingdom, and Yugoslavia. Austria has since joined, and Yugoslavia withdrawn. In addition, Spain was a member during the 1960s and rejoined in 1983.

Research Policy 13 (1984) 183–210
North-Holland

0048-7333/84/$3.00 © 1984, Elsevier Science Publishers B.V. (North-Holland)

Européen pour la Recherche Nucléaire – a preparatory body set up in 1952 – ceased to exist, although the new organization retained the acronym "CERN." The main purpose of CERN was outlined in Article II of the Convention, and is quoted above.

By 1954, the interim Council had already determined the main components of the future programme for CERN, in particular the construction of a 600 MeV (million electron-volts) synchro-cyclotron, and "a proton synchrotron for energies above ten giga electron-volts" (see e.g. [3, p. 7]). Today, thirty years on, the early estimates made about the future cost and size of the laboratory seem surprisingly modest. In 1952, for example, it was

> estimated that the next stage of development will last seven years and cost a total of 27 million dollars after which the estimated annual running cost of the laboratory would be a maximum of $1\frac{1}{2}$ million dollars [22, p. 1105].

As for the number of staff, it was calculated that "a total staff of about a hundred scientists, draughtsmen and technicians" [23, p. 647] would be sufficient to build the Proton Synchrotron (PS). However, by the end of the initial seven-year development programme in 1960, the total number of CERN staff was well over 1000, and expenditure for that year (the first after the completion of the PS) was also approximately ten times the early estimates – 66 million Swiss francs or about 15 million dollars. CERN continued to grow rapidly, reaching a peak of nearly 3800 employees in 1975. It has since decreased a little in size, so that in 1982 it employed approximately 3600 staff and had an annual budget of over 600 million Swiss francs (MSF) (around 300 million US dollars). As for the new accelerator, LEP, it is estimated that Phase 1 of this project (to reach an energy of 50 GeV in each of two colliding beams) will cost 910 MSF (nearly 450 million US dollars), while the cost of Phase II (to reach 130 GeV) has yet to be publicly specified.

In view of this considerable long-term investment of scientific and financial resources in CERN, it is pertinent to ask how successful the CERN accelerators and the high-energy physicists who use them have proved. As a prominent theoretical high-energy physicist recently remarked in connection with CERN, "the tax-payer is entitled to know what he is getting for his money" (Polkinghorne [30]). But what criteria should be adopted in judging the success of the CERN accelerators and their users? A few years ago, a senior CERN official gave the following explicit definition of the criterion to be used:

> CERN must in the end be judged on whether it has succeeded in its primary mission of enabling particle physics in Europe to be maintained at a world level (Hine [9, p. 181]).

How highly do the scientific outputs associated with the CERN accelerators rank in world terms? At CERN itself, there is little doubt about the answer to this question, as the following quotations make clear.

> The main purpose of CERN is to provide Europe's scientists with first class facilities for high-energy physics research … The output is measured in new knowledge and in trained people. On these criteria, CERN can stand comparison with any laboratory in the world … [4, p. 111].

> CERN was conceived to help restore the quality of European science, to provide research facilities beyond the means of individual countries … In the past twenty-five years, these hopes have been fulfilled beyond the expectations of any of CERN's creators. From accelerators and storage rings, Europe's scientists have contributed greatly to our knowledge of the nature of matter. The CERN laboratory … puts European high-energy physics on a par with that of any other region of the world [5, p. 227].

But can we go beyond these general statements to establish just how "great" have been the contributions from the CERN accelerators [2] to the advance of scientific knowledge? Where precisely do the CERN experimental contributions [3] rank in relation to those from other accelerator laboratories? It is the purpose of this paper to attempt to answer such questions. We begin with a brief examination of the history of CERN.

---

[2] It should be stressed that what is being assessed in this paper and paper II [12] is the scientific performance of the CERN accelerators and of the high-energy physicists who have used them, rather than CERN (the laboratory and its staff) *per se*.

[3] We have not considered here the theoretical outputs from CERN, nor have we examined other types of contribution from the CERN accelerators, such as their training function. For example, the good design and rapid success of the recent proton–antiproton collider experiments (discussed in paper II [12]) clearly owe much to experience with the ISR where physicists took some time to appreciate the importance of 4π-solid angle detectors.

2. The history of CERN and its main accelerators

In 1950, UNESCO instructed its Director-General to encourage the formation of international research centres to increase co-operation in fields where the costs were too great to be borne by most individual countries. High-energy physics (or nuclear physics, as it was still then known) was identified as one such field. While accelerators similar in energy to the 3 GeV Cosmotron and 6 GeV Bevatron, then being built at the Brookhaven and Lawrence Radiation Laboratories in the United States, might just be within the reach of individual nations (and several accelerators of such an energy were eventually built), to go beyond these energies was seen to necessitate international collaboration. This was the principle underlying the multinational collaboration at CERN that over the years was to result in the construction of a series of major particle accelerators.[4]

Initially, work began at CERN in 1952 on plans for the construction of a 10 GeV machine. However, after the design-team had visited Brookhaven, where a similar design-study was in progress, they concluded that, by incorporating the newly invented "strong focusing" concept, it would be possible to built a 30 GeV accelerator for the same cost as the lower-energy "weak focusing" machine originally planned. The interim Council agreed to this change, although with the proviso that the maximum energy be limited to 25 GeV.[5] This was several GeV less than the design-energy of the Brookhaven accelerator, a difference which was subsequently to leave the CERN PS at a significant disadvantage for certain experiments; but otherwise the two designs were very similar. Indeed, although a certain element of competition existed between the CERN and Brookhaven teams during the design and construction period for the two accelerators, a considerable and very fruitful interchange of ideas and technical information also took place (cf. Johnsen [14, pp. 37–39]).

Meanwhile, the design and construction of the smaller 600 MeV synchro-cyclotron went ahead, the first protons being accelerated in 1957. The synchro-cyclotron is still operating over twenty-five years later. During this time, it has supported a wide programme of research, initially in particle physics but increasingly in nuclear physics. However, because such research is no longer classified as "high-energy physics" (which is now normally defined as "physics done with accelerators able to produce primary particles at an energy higher than 1 GeV" – see, for example, Roche [31, p. 3]), and because the costs involved have represented only a small fraction of the CERN annual budget (3 percent in 1977 – see CERN [2, p. 25]), the synchro-cyclotron will not be further considered here.

The Proton Synchrotron (PS) was eventually completed in late 1959, a few months ahead of schedule and – perhaps more importantly from the point of view of CERN's self-confidence – before the rival Brookhaven accelerator. Its maximum energy, at 28 GeV, was also a little higher than the CERN Council had suggested. The PS has been in operation ever since, although it has been subject to a continuous programme of modification and improvement with the result that its intensity (the number of protons accelerated per pulse) in the latter stages of its lifetime has been 1000 times greater than the original design figure.

Before considering subsequent developments at CERN, it is necessary to examine briefly the structure of experimental high-energy physics in Western Europe at the start of the 1960s. By this time, accelerators of between 1 and 3 GeV were already operating in Britain at Birmingham University, at Saclay and Orsay Laboratories in France, at the Frascati Laboratory in Italy, and at Lund University in Sweden. In addition, various accelerators were under design or construction at Rutherford Laboratory and, a little later, at Daresbury Laboratory in the UK, at DESY, the German laboratory near Hamburg, and at the University of Bonn. V.F. Weisskopf, then Director-General of CERN, clearly saw the need to co-ordinate national and international efforts in Western Europe and to plan for the longer term. Largely at his initiative, an informal committee was set up in 1963 to make recommendations about the future development of high-energy physics in Europe.[6]

---

[4] For further details of the early history of CERN, see Goldsmith and Shaw [8]. In addition, a major historical study of CERN is currently being undertaken by a multinational team under the direction of Professor A. Hermann.

[5] According to one of those involved, the Council insisted upon this because of their worry that over-ambition on the part of the design-team might lead them to attempt too demanding a project (Interview, 1981).

[6] This was known as the Amaldi Committee and reported to the CERN Council. For details, see Amaldi [1].

This committee, which subsequently came to assume a more permanent basis as ECFA, the European Committee for Future Accelerators, reported later in 1963. Its two main recommendations were that Europe should build a 300 GeV proton accelerator, and that an intersecting storage rings facility should be added to the Proton Synchrotron at CERN. In addition, the Committee argued that, for "a minimum balanced programme of research in high-energy physics," it was essential that these major European machines be supported by a pyramid of national and regional accelerators of lesser energy. It was envisaged that altogether these would cost up to four times as much as the 300 GeV machine itself (see e.g. Amaldi [1]). Not surprisingly,

> a good many custodians of public funds took fright and thought they were being asked not merely to build the big machine but to provide high-energy physics with a blank cheque [24, p. 1283].

To those nations which had joined CERN in the expectation that the construction of accelerators on an international basis would be cheaper than each country attempting to build its own, the argument that such international facilities in fact required for their full exploitation a very considerable increase in expenditure on national facilities, must have come as something of a surprise. Five years later in 1968, when Britain initially decided against participating in the 300 GeV project, the effects of this report were evidently still being felt, as this commentator's observation makes clear:

> The first attempt to win public support for the big accelerator made quite unrealistic demands that more spending at CERN should be accompanied by more spending domestically, and this gaffe has never been openly repudiated [25, p. 1197].

Despite these question-marks over the wisdom of requesting funding for up to nine additional national and regional accelerators to form the "base of the pyramid programme" (Amaldi [1, p. 1290]), preparations for the two facilities planned to occupy the summit of the pyramid went ahead. The idea of storing protons and causing them to collide head-on in two intersecting rings had been taken up at CERN in 1960, building upon ideas developing by D. Kerst and G. O'Neill in the United States and G. Budker in the Soviet Union. [7]

After successful testing of these ideas in 1963 on an electron model (CESAR) built at CERN, the Intersecting Storage Rings (ISR) project was approved by the CERN Council in 1965. One of the evident attractions of the project was its uniqueness – for the first time since the War, Europe was building a major accelerator for which there was no American counterpart. V.F. Weisskopf, Director-General of CERN in the crucial early years of the ISR discussions, lent considerable support to the project because of his

> deep conviction that the physicists of Europe can, and should, be not only on a par with other scientific communities, but ahead at least in some aspects (Weisskopf [35, p. 291]).

Despite the novelty of the machine and the formidable technical problems that it posed, the ISR was completed on schedule – the world's first proton-proton interactions in colliding beams being observed early in 1971. The ISR was in regular operation until the end of 1983, and, like the PS, was considerably modified and improved. In particular, its luminosity (the quantity that, when multiplied by the cross-section in cm^2 for a given process, yields the event rate per second) was increased from an initial value of $10^{28}\ \mathrm{cm}^{-2}\ \mathrm{s}^{-1}$ in 1971 to 4×10^{30} (the design figure) by the end of 1974, and to 1.2×10^{32} by March 1981, an improvement of over four orders of magnitude.

Before going on to discuss the Super Proton Synchrotron, mention should be made of one other important milestone in the history of CERN – the signing in 1967 of an Agreement with the USSR State Committee for the Utilization of Atomic Energy under which a joint scientific and technical programme would be mounted at the 70 GeV proton synchrotron then nearing completion at the Serpukhov Institute of High-Energy Physics. In return for CERN providing technical help and equipment (for example, a rapid beam-extractor) for the Soviet machine, West European physicists were granted access to what was, for a few years, the highest-energy accelerator in the world. [8]. The Agreement was extended in 1975 to allow Soviet teams to participate in experiments at CERN, and again in 1983 to cover new machines at CERN and Serpukhov.

---

[7] However, the first idea of using head-on collisions between particles was apparently due to Wideroe in 1943, an idea he subsequently patented in 1953.

[8] An evaluation of the past scientific performance of the Soviet accelerator at Serpukhov is made in Irvine and Martin [11].

The first design for a 300 GeV Super Proton Synchrotron (SPS) was drawn up in 1964, but its construction was not finally authorized until 1971. One factor underlying the delay was the cost of the original design – in 1967, a figure of 1800 MSF (about $420 million at the exchange rates then prevailing) was quoted (Amaldi [1, p. 1290]). This was considerably more than the estimated $240 million cost of an equivalent 200–400 GeV accelerator at Fermi National Accelerator Laboratory, Batavia, Illinois, the design of which was completed in 1967 under the direction of R.R. Wilson. A second reason for the delay was Britain's initial withdrawal from the project in 1968.[*] Since it was intended that Britain contribute nearly one quarter of the cost, this was potentially a severe blow; however, almost immediately CERN was able to propose a revised project costing 25 percent less.

Third, there was profound disagreement among CERN Member States over the site for the accelerator, with several arguing strenuously that it should be built on their soil. Eventually in 1970, a compromise was reached whereby the SPS would be built at Geneva – thus avoiding the political difficulties of agreeing a new site – using the existing Proton Synchrotron as an injector and further reducing the costs to just over 1100 MSF. Moreover, 200 MSF of this total would now be found from reductions in CERN's existing commitments. Hence, only 900 MSF of "new money" would be required, half the cost of the original design and low enough to encourage Britain to rejoin the project. This dramatic reduction of the original cost estimate gave rise to a feeling in the scientific community that CERN

> has often been too presumptuous in its belief that Europe owes high-energy physics a living [26, p. 593]

and that

> if it had not been for the shining example of Dr. R.R. Wilson's bigger but cheaper accelerator at Batavia, Illinois, the chances are that Europe would still be lumbered with a needlessly expensive machine [27, p. 1014].

The SPS proposal was approved in 1971, four years after authorization for the very similar machine at Batavia. This gap was to be maintained, with commissioning of the SPS coming four years after the US accelerator became operational. The effects of this significant lead are discussed in paper II (Irvine and Martin [12]) which considers the relative scientific outputs of these two machines.

The SPS operated continuously until mid-1980 when it was closed for a year to permit construction of an associated proton–antiproton collider. The idea of using the SPS to collide protons and antiprotons had been put forward in 1976 by C. Rubbia, and, following successful tests of the crucial technique of stochastic beam-cooling (developed by S. van der Meer of CERN), the project was approved in 1978. It transpired that the SPS, and in particular its vacuum system, had been so well designed that it could be quickly and relatively cheaply converted into a proton–antiproton storage ring. In contrast, the rival Fermilab machine required extensive structural modifications to add a similar, though higher-energy, storage-ring capability. As a result, it has taken several years longer to convert. The main justification for both these machines was to open up a new energy-region where several phenomena underlying the currently accepted "standard theory" (which unifies electromagnetic and weak interactions) were expected to occur, and in particular to look for the "intermediate vector bosons" believed to be the carriers of the so-called weak force. Because such a discovery would be a major achievement in the history of high-energy physics, there was inevitably a certain element of a "race" between rival laboratories around the world to find these new particles first. This race was "won" by the users of the CERN collider, who early in 1983 announced the discovery of the charged W particle and a few months later the neutral Z particle. These discoveries came after our study had been completed, but they are discussed in paper II.

Finally, brief mention should be made of LEP, although detailed analysis of the prospects for this accelerator are left until paper III (Martin and Irvine [20]). During the 1970s, CERN considered several options for a major new machine for the 1980s and 1990s, and, in the end, after extensive discussions within the European high-energy physics community, settled on the choice of a large electron–positron collider (LEP) capable of producing beams with an energy initially of 50 GeV (Phase I), and eventually of up to 130 GeV. The construction of LEP was authorized at the end of 1981, and the first collisions are planned for 1988.

---

[*] For a full analysis of the political and scientific factors underlying this decision, see Gibbons [7].

This concludes our discussion of the main developments in CERN's history. It should be noted that, during this time, the structure of West European high-energy physics has changed considerably. Whereas in 1970, there were some nine national and university accelerators of 1 GeV or more, providing research facilities for perhaps half of Europe's experimental high-energy physicists, now those experimentalists are largely concentrated in just two centres – CERN and DESY (the German facility where international participation has greatly increased with the construction of first DORIS, and then PETRA, two electron–positron colliders [10]).

A similar trend is apparent in the United States: in 1970, besides the three large laboratories at Brookhaven, Fermilab, and Stanford, experimental high-energy physics was also carried out at Argonne, Berkeley, Cambridge, Cornell, and Princeton. Of these latter five, only Cornell is still in the field, with further rationalization a distinct possibility. [11] In view of the growing concentration of Western Europe's high-energy physics effort on the CERN accelerators (nearly three-quarters of the 2000 European experimentalists currently rely on these for their research), it has obviously become an important policy issue to devise adequate methods of assessing the past scientific performance of such facilities, particularly in relation to the main competitor accelerators at the three large United States laboratories.

3. The assessment of scientific performance

How can the performance of experimental high-energy physics facilities be assessed and compared? In considering this question, it should be noted that high-energy physics is funded primarily for scientific reasons – i.e. for the contributions it is judged likely to make to the advance of knowledge. Experimental high-energy physics activity does undoubtedly also lead to significant educational benefits (in the form of highly-trained scientific and engineering personnel), to various types of technological "spin-off" (see Schmied [32]), and, in the case of multinational organizations like CERN, to the fostering of international co-operation. But the primary reason why Member States are prepared to contribute over 600 MSF a year to CERN is because of the promise of important scientific results in coming years. Judgements on the likely future performance of CERN will be based in part on the research potential of programmes on existing and planned accelerators, and also on the "track-record" of scientists in exploiting its facilities to contribute to scientific progress. How can such contributions to scientific progress be reliably assessed?

In a previous study, we developed the method of "converging partial indicators" in which a number of (partial) indicators of scientific progress (publication counts, citations, numbers of highly cited papers or "discoveries," and extensive peer-evaluation) were used to assess matched groups of researchers using similar research facilities in a given field (see Martin and Irvine [19] for details). This method enables assessments to be made of the *productivity* of these facilities and their associated user-groups, the *impact* of the research results, and the *perceived significance* of the research within the scientific community, results which can be combined to give an overall evaluation of their contribution to "scientific progress." It was applied in three different Big Science specialities – radio astronomy, optical astronomy, and electron high-energy physics (see Martin and Irvine [19]; Irvine and Martin [10]; and Martin and Irvine [18]). In each case, a certain convergence between the results based on the various partial indicators was obtained, which permitted conclusions to be drawn as to the relative contributions to scientific progress associated with each of the experimental facilities. The main elements of the method are summarized in table 1. [12]

---

[10] See Martin and Irvine [18] for an assessment of the earlier DESY machines.

[11] See, in particular, the report of the recent US Department of Energy Subpanel on Long-Range Planning for High-Energy Physics, chaired by G. Trilling [34].

[12] In view of the comments of scientists critical of the value of citation analysis, it is worth re-emphasizing that citations are used in our method of research evaluation to indicate the "impact" of publications on the advance of knowledge, but not necessarily their "quality" or "importance." Thus a high quality paper in a stagnating field may contribute little to the advance of knowledge in general, and is likely to receive fewer citations than a similar quality paper in a more active field. The "importance" of a paper is defined as the influence it would have were scientific communication completely free from institutional, social and political constraints. For example, a potentially influential paper may go

Table 1

Main problems with the various partial indicators of scientific progress and details of how their effects may be minimized using the method of converging partial indicators

Partial indicator based on	Problem	How effects may be minimized	
A. Publication counts	1. Each publication does not make an equal contribution to scientific knowledge	Use citations to indicate average impact of a research facility's publications, and to identify very highly cited papers	Use only indicators that yield convergent results
	2. Variation of publication rates with specialty, institutional context, etc.	Choose matched research facilities producing similar types of papers within a single specialty	
B. Citation analysis	1. Technical limitations with *Science Citation Index*:		
	(a) first author only listed	Not a problem when dealing with a research facility	
	(b) variations in names (c) authors with identical names (d) clerical errors	Check manually	
	(e) incomplete coverage of journals	Not a serious problem for Big Science	
	2. Variation of citation rate during lifetime of a paper – unrecognized advances on the one hand, and integration of basic ideas on the other 3. Critical citations 4. "Halo-effect" citations	Not a problem if citations regarded as an indicator of impact, rather than quality or importance	
	5. Variations of citation rate with type of paper and specialty	Choose matched research facilities producing similar types of papers within a single specialty	
	6. Self-citation and "in-house" citation (SC and IHC)	Check empirically and adjust results if the incidence of SC or IHC varies between groups	
C. Peer evaluation	1. Perceived implication of results for own research facility and competitors may affect evaluation 2. Individuals evaluate scientific contributions in relation to their own (very different) cognitive and social locations 3. "Conformist" assessments (e.g. "halo effect") accentuated by lack of knowledge on contributions of different research facilities	1. Use a large representative sample 2. Use verbal rather than written survey in order to press evaluator if a divergence between expressed opinions and actual views is suspected 3. Assure evaluators of confidentiality 4. Check for systematic variations between different groups of evaluators	

In what follows, we apply this method to what have been the three principal CERN facilities over the last two decades – the Proton Synchrotron (PS), the Intersecting Storage Rings (ISR), and the Super Proton Synchrotron (SPS) – comparing them with their main competitors elsewhere in the world, in particular the Brookhaven Alternating Gradient Synchrotron (AGS) on the United States East Coast, the Serpukhov 70 GeV proton synchrotron in the Soviet Union, and the Fermilab 400 GeV proton synchrotron in the American Mid-West. It should be stressed from the outset that our aim is not so much to assess CERN *per se*, but *the scientific output from the CERN accelerators*, irrespective of whether they were used by CERN staff, user-groups from European Member States, or visiting scientists from the United States or Eastern Europe. We are thus comparing the overall contributions to scientific progress made using CERN accelerators with those made using other accelerators. We are *not* assessing the performance of CERN staff (some of whom, for example, were involved in a successful series of experiments at Serpukhov), nor even the performance of West European high-energy physicists – although 80–90 percent of CERN users come from the 13 Member States, some of the major experimental advances (particularly with the ISR) have involved United States groups. It is also important to note the crucial distinction between (1) the overall scientific output from each accelerator, irrespective of the scale of funding and number of users; and (2) their scientific "productivity" or output per unit of input, i.e. their output in relation to funding and numbers of users.

Some comment should also be made concerning the time-period considered in the present study. The first research papers from the CERN PS were published in 1960, with the result that an assessment should, if possible, cover the period from then to the present. This requirement, however, raised a problem for our methodology, since peer-evaluation, which provides perhaps the least problematic of the various partial indicators of scientific progress, only works well for a period of up to ten years or so. There are two reasons for this: memories tend to fade rather quickly outside this time-scale, and, in a scientific field like high-energy physics, many of those involved have only been actively engaged in research for ten years or less.[13]

The approach adopted in our assessment, therefore, is as follows. First, we focus on a recent ten-year period in which peer-evaluation data can be used to complement the full range of other indicators. The period chosen was from 1969 to 1978, partly because this is compatible with that used in an earlier study of electron accelerators (Martin and Irvine [18]), and partly because scientists tend to find it difficult to assess reliably the significance of very recent work. For this period, statistics at a fairly high level of aggregation are used; for example, the three CERN accelerators are treated as one unit. We can then examine the degree of convergence between the quantitative peer-evaluation data and the indicators based on bibliometric (publication and citation) data, the results of which are described in the remainder of this paper. The second time-frame considered covers the 22 years from 1961 to 1982. Given the problem of obtaining systematic peer-evaluation data over such an extended period, we base the assessment largely on the bibliometric indicators interpreted with the aid of qualitative information obtained during interviews with high-energy physicists, together with a much more limited amount of quantitative peer-evaluation data. The data used in this second comparison, which forms the subject matter of paper II, are presented in a more disaggregated form so that each CERN accelerator can be considered separately, and changes over time studied in greater detail. An attempt is then made to interpret the differences in scientific performance of the CERN accelerators relative to their closest competitors, and to identify possible factors explaining those differences.

---

unnoticed and uncited if it is written in an obscure or non-English language journal. The "impact" of a publication, in contrast, describes its actual influence on the advance of knowledge, and it is this for which citations provide a (partial) indicator. Note that a "mistaken" paper can sometimes have a significant impact on the advance of knowledge if it stimulates research that might not otherwise have been carried out, and its citation record may reflect this (see Martin and Irvine [19, p. 71] for further discussion).

[13] Twenty-five percent of the 182 researchers interviewed during the study received their Ph. D. in 1970 or later.

4. Inputs to high-energy physics

It is necessary to examine in some detail the various inputs to experimental high-energy physics, in particular the numbers of users for the accelerators, and their respective scales of financial support.[14] This provides the information required below to arrive at judgements on the relative scientific "productivity" (i.e. scientific outputs per unit of input) of different accelerators.

For Western Europe and the United States we have drawn heavily upon comparative data produced by C. Roche, Head of the Central Planning Office at CERN, and by W. Kirk of the Stanford Linear Accelerator Center (SLAC) in California. Roche has, in turn, used information from censuses conducted by ECFA, the European Committee for Future Accelerators, while Kirk has access to the relevant Department of Energy and National Science Foundation data for the United States. Supplementary information comes from the laboratories themselves – for example, *CERN Annual Reports* give data on numbers of users and annual expenditure in previous years – and this has been used to check the accuracy of the figures. This cross-checking suggests the European and United States figures are accurate to within 10 percent or so.

Data on the approximate numbers of users of accelerators at CERN and DESY in Europe, and at the three principal US laboratories, are shown in table 2 for the years 1970, 1974, and 1978. It is important to note that graduate research students have been included in our figures. This has not always been the case in previous studies of the size of the high-energy physics community, particularly in the United States, with the result that some of the previously quoted figures have been rather smaller.[15] We have then attempted to estimate the total number of high-energy physics experimentalists in the world, in order to calculate the approximate percentage of active researchers using the accelerators at each of the five largest Western laboratories. The figures reveal that, for CERN, the figure has risen from just under 20 percent in 1970 to approximately 30 percent in the second half of the 1970s, as smaller national accelerators were closed or diversified from high-energy physics to nuclear physics and synchrotron-radiation research. It should be noted, however, that the CERN figure for 1970 does not include the 100 or more European physicists preparing experiments to be run on the ISR in later years – these are part of the figure of 1000 quoted for "other" West European experimentalists. In the same way, the 1974 figure of 550 "others" in Western Europe includes some in the early stages of planning experiments for the CERN SPS. Similar considerations apply to our data on the users of accelerators at the other laboratories.[16] Despite these qualifications, it can nevertheless be seen that during the 1970s the CERN accelerators have catered for a far larger number of physicists than Fermilab (used by 11–12 percent of world experimentalists), Brookhaven and SLAC (both approximately 5–6 percent), or DESY (whose percentage share has risen from 3 percent to over 6 percent, as first DORIS and then PETRA, the two electron–positron storage rings, came into operation).

The production of reliable data on which to base international comparisons of funding is even more difficult. There are several major problems here. First, the figures must be fully comparable – that is, they must be based on common categories and definitions. In view of the importance of such figures in evaluations of a centre's relative cost-effectiveness, this has, in the past, given rise to some debate between C. Roche at CERN and W. Kirk at SLAC. They have, however, now succeeded in reaching approximate agreement on the criteria to be used in such comparisons, and, as far as possible, we have adopted their categories. Second, one must allow for the varying inflation rates in different countries. Third, fluctuations in exchange rates

---

[14] In previous papers, we have considered other inputs, such as numbers of technical support staff, as well as attempting to separate capital costs from recurrent expenditure. However, this gives little additional information, and we assume here that numbers of users and overall funding levels (capital and recurrent) are sufficient to define the relative scale of research activity at each accelerator (with certain provisos discussed later).

[15] For example, see Moravcsik [21], p. 84]. We include graduate research students in these comparisons since they form an integral part of most experimental high-energy physics teams, for example often undertaking complex data-processing tasks.

[16] In addition, the figure shown in table 2 for SLAC in 1978 is smaller than that quoted in Martin and Irvine [18, table 1] because the latter included a large number of experimentalists preparing to carry out future experiments on the new electron–positron collider, PEP.

Table 2
Numbers of experimental high-energy physicists (HEPs)[a]

		1970	1974	1978
W. European HEPs	CERN users	850 (18%)	1250 (28%)	1400 (32%)
	DESY users	150 (3%)	200 (4.5%)	280 (6.5%)
	Others[b]	1000 (21%)	550 (12%)	320 (7.5%)
	Total W. Europe	2000 (43%)	2000 (44%)	2000 (45%)
US HEPs	Brookhaven users	300 (6.5%)	~ 250?[d] (5.5%)	200 (4.5%)
	Fermilab users	–	~ 450?[d] (11%)	550 (12%)
	SLAC users	200 (4.5%)	300 (6.5%)	250 (5.5%)
	Others[b]	1000 (21%)	~ 300? (5.5%)	200 (4.5%)
	Total US	1500 (32%)	1300 (29%)	1200 (27%)
HEPs in rest of world (E. Europe, Japan, etc.) – estimate[c]		700–1700?	700–1700?	700–1700?
Estimated world total		4700(± 500) (100%)	4500(± 500) (100%)	4400(± 500) (100%)

[a] These include Ph. D. students carrying out research work. As with funding, these are approximate figures only. The data for 1978 are probably the most accurate, the European figures having been taken from Roche [31, p. 5]. Other data came from annual reports (e.g. CERN [3, p. 17], internal documents provided to us by the various laboratories, and Kirk [15, fig. 5].

[b] These include the users of various smaller accelerators and of accelerators overseas, as well as a number of physicists in the very early stages of planning new experiments or analyzing experimental data obtained in previous years.

[c] These are very approximate estimates only. It is assumed that the effect of the decreasing number of Soviet accelerators used for high-energy physics has been largely compensated for by the growth of the subject in other countries such as Japan.

[d] Estimates only.

must also be taken into account. Lastly, there is the problem that official exchange rates do not necessarily reflect actual purchasing power in different countries. To overcome these second and third problems, we have used the two specialized inflation indices (based on changes in the cost of scientific equipment, salaries, etc.) produced by CERN and SLAC[17] to convert all expenditure figures to 1978 prices (in Swiss francs for Europe, and dollars for the US). Then, to overcome the fourth problem, the Stanford Research Institute "competitive value index" – i.e. the exchange rate corrected for the real purchasing power of the dollar in the United States – has been used to convert all the US figures (in 1978 prices) to Swiss francs. For 1978, the "competitive value index" was $1 = 2.7 Swiss francs, an exchange rate that both Roche [31, p. 4] and Kirk [15, The Richter Method] take to be realistic to within 10 percent or so.

Table 3 gives the estimated figures for overall European and US government spending on high-energy physics in both actual prices and equivalent 1978 prices, while in table 4 the US figures have been converted from 1978 dollars to 1978 Swiss francs. Table 4 also contains a very approximate estimate of total world expenditure on high-energy physics, which has been used to convert the figures for the major West European and US facilities into percentage shares of total expenditure. These figures suggest that, in the case of CERN, this percentage has grown from 15 percent in 1970 to some 25 percent in the second half of the 1970s. In addition, a large fraction of the "other" West

---

[17] It has been assumed that the SLAC inflation index is appropriate for the other American laboratories and that the CERN index is not too dissimilar to that for DESY.

Table 3
European and US funding for high-energy physics (HEP)[a]

		1966[b]		1970[b]		1974[b]		1978[b]
		1966 prices	(1978 prices)[c]	1970 prices	(1978 prices)[c]	1974 prices	(1978 prices)[c]	1978 prices
W. European funding in millions of Swiss francs (MSF)	CERN[d]	165	(270)	302	(440)	580	(670)	565
	DESY	50	(80)	70	(100)	100	(115)	115
	Other	335	(550)	430	(620)	480	(555)	375
	Total W. Europe	550	(900)	802	(1160)	1160	(1340)	1055
US funding[e] in millions of dollars ($M)	Brookhaven	26	(57)	32	(59)	33	(48)	42
	Fermilab	–	(–)	56	(102)	65	(94)	84
	SLAC	29	(64)	27	(49)	26	(38)	87
	Other	109	(240)	117	(215)	63	(91)	90
	Total US	164	(361)	232	(425)	187	(271)	303

[a] These are the estimated figures for total expenditure on high-energy physics (i.e. capital and construction costs as well as recurrent expenditure).

[b] The figures are for calendar years in W. Europe, and for fiscal years in the US.

[c] Inflation indices calculated by C. Roche (at CERN) and W. Kirk (at SLAC) have been used (see Kirk [15, fig. 3]. The conversion factors for changing the 1966 figures into equivalent 1978 prices are 1.64 (W. Europe) and 2.20 (US); for 1970, 1.45 and 1.83, and for 1974, 1.15 and 1.45.

[d] These are the figures for the contributions from Member States to CERN (i.e. they exclude bank interest, miscellaneous income, etc.). In addition, an allowance for non-HEP programmes at CERN has been made by subtracting a figure of between 5 percent (in 1966) and $3\frac{1}{2}$ percent (in 1978).

[e] These figures are based on (1) data provided by the US Department of Energy (DoE) which show DoE support of $148 M in 1966, $215 M in 1970, $169 M in 1974, and $280 M in 1978, and (2) data quoted in Kirk [15, fig. 3] showing National Science Foundation support of $16 M, $17 M, $18 M and $23 M in 1966, 1970, 1974 and 1978 respectively.

Table 4
World funding for high-energy physics (1978 Swiss francs)

		1966 MSF[a]	1970 MSF	1974 MSF	1978 MSF
W. European funding	CERN	270 (10%)	440 (15%)	670 (26%)	565 (24%)
	DESY	80 (3%)	100 (3.5%)	115 (4.5%)	115 (5%)
	Other	550 (22%)	620 (21%)	555 (21%)	375 (16%)
	Total W. Europe	900 (35%)	1160 (39%)	1340 (52%)	1055 (44%)
US funding[b]	Brookhaven	145 (5.5%)	160 (5.5%)	130 (5%)	115 (5%)
	Fermilab	–	275 (9%)	255 (10%)	225 (9.5%)
	SLAC	175 (6.5%)	135 (4.5%)	100 (4%)	235 (10%)
	Other	650 (25%)	580 (19%)	245 (9.5%)	245 (10%)
	Total US	970 (37%)	1150 (38%)	730 (28%)	820 (34%)
Estimated funding[c] – rest of world (E. Europe, Japan, etc.)		~ 500–1000?	~ 500–900?	~ 400–700?	~ 400–700?
Estimated world total (billions of Swiss francs)		2.6 BSF (± 0.3) (100%)	3.0 BSF (± 0.2) (100%)	2.6 BSF (± 0.2) (100%)	2.4 BSF (± 0.2) (100%)

[a] Millions of Swiss francs.

[b] A conversion rate of 1$ = 2.7 SF in 1978 has been used. This is based not on the official exchange rate, but on the Stanford Research Institute "Competitive Value Index" by which exchange rates are corrected for real purchasing power in different countries (see Kirk [15]; Roche [31, p. 4]).

[c] These are very approximate estimates only. In deriving them, it has been assumed that countries outside Western Europe and the United States spend somewhere between 50 and 100 percent of the amount budgeted by the United States. In 1966, when the Soviet Union was constructing the large accelerator at Serpukhov, the figure is likely to have been in the upper part of this range, while in 1970 (when the US was in the midst of building the Fermilab accelerator) it is more likely to have been in the lower part.

European funds is spent on supporting university groups using the CERN accelerators, a fraction that has undoubtedly been increasing during the 1970s as national accelerators were closed and the percentage of European experimentalists using CERN rose from less than 45 percent in 1970 to 70 percent in 1978. This change in the structure of funding can be taken into account by allocating the "other" funds in the ratio of CERN users to other European experimentalists (including DESY users), having first subtracted a small amount to cover the cost of the few remaining national facilities. The result is that in 1978, for example, the CERN share of world expenditure is raised from 24 percent to over 30 percent, bringing it closely into line with the figure in table 2 on CERN users as a percentage of the world total of experimentalists. (In 1978, CERN accelerators catered for 32 percent of all experimentalists.)

Similarly, the US "other" funds were utilized partly to support university users of Brookhaven, Fermilab, and SLAC, and partly to operate the Argonne and Cornell accelerators in 1978, and the Berkeley, Cambridge and Princeton accelerators in earlier years. Thus, apportioning part of the US other costs to the three national laboratories raises the shares of world expenditure of Fermilab to 12–13 percent, Brookhaven to 6–7 percent, and SLAC to about 6 percent (except in 1978, when it was much higher because of the construction of PEP), figures again very close to the user-percentages given in table 2.

The results of integrating these "other" expenditures for both Europe and the United States in the way described are shown in table 5, alongside figures for the distribution of world experimentalists over the ten-year period 1969–78. The close similarity of these figures for all five centres suggests that the cost per experimental high-energy physicist does not differ greatly between centres (cf. Roche [31, table 1 and p. 6]). This is provided that (a) one corrects official exchange rates for real purchasing power; and (b) it is recognized that the total expenditure in a given year may be untypical because of a large capital-construction element. Thus, the figures in table 6 on the approximate cost per high-energy experimentalist seem to show higher costs for America in 1970 and 1978, and for Europe in 1974. Yet these apparent differences are very largely due to the fact that the US figure for 1970 contained a significant element for the construction of Fermilab, and that for 1978 included the construction costs of PEP and CESR, the Cornell Electron Storage Ring, while the European figure for 1974 was swollen by the construction costs of the SPS. With these reservations, the figures in table 5 can be regarded as giving a reasonable indication of the scale of inputs to the CERN accelerators relative to those for other major facilities.

Table 5
Inputs for major accelerators at the five main Western high-energy physics centres, 1969–78

	Users (% of world total)	All-inclusive funding (% of world total)
CERN	~ 25–32 [a]	~ 25–32 [a]
DESY	~ 4 [b]	~ 4–5 [b]
Brookhaven	~ 5–6	~ 6–7
Fermilab	~ 11–12 [a]	~ 12–13
SLAC	~ 5–6	~ 6 [b]

[a] Except in the very early years of the decade, 1969–78.
[b] Except in the last year or so of the decade.

Table 6
Approximate cost per experimental high-energy physicist [a] (in 1978 Swiss francs)

	1970 MSF	1974 MSF	1978 MSF
W. Europe	0.58	0.67	0.53
United States	0.77	0.56	0.68

[a] These data were obtained by dividing overall funding for high-energy physics (adjusted to 1978 prices – see table 4) by the number of experimentalists. This procedure ignores the fact that some funds are used to support theoretical work. However, such research is relatively cheap compared with experimental work (at CERN in 1977, for example, it accounted for only 1½ percent of the total budget – see [2, p. 25]) so the effect of this approximation should not be too great.

5. Accelerator outputs – scientific publications

While publication counts in themselves constitute a rather problematic indicator of scientific progress, they nevertheless provide some useful information and are a necessary complement to the other partial indicators used in this study. The first step in producing these figures was to compile as comprehensive a list as possible of all experimental high-energy physics papers published over the period 1961–82. This was done by scanning fully the 11 principal international journals [18] used by particle physicists, noting details of all papers which would generally be regarded as coming under the category of "experimental high-energy physics". [19] Each paper was then read to establish which accelerator (or accelerators [20]) had been used to produce the experimental data reported. It should be noted that, under this procedure, preprints and conference papers were excluded from consideration on the grounds that nearly all such articles are eventually published in a scientific journal, and to have included them would have introduced an element of "double-counting". [21]

The list thus derived was then cross-checked against various data compilations (especially Particle Data Group [28]) and annual publication lists provided by the main centres. [22] This increased the total number of publications by about 5 percent. Some of these additional papers involved borderline definitions as to whether the results reported were "experimental," or indeed whether they should be treated as "high-energy physics" – if there was any doubt, they were included. The remaining additional papers were published in journals other than the 11 scanned, mainly in general physics, or "national" high-energy physics, journals. Cross-checking in this way suggested that our final publication list was 90–95 percent complete in its coverage of papers from Western accelerators, but rather less in the case of Soviet and Japanese accelerators. This should be borne in mind when considering the figures in table 7, and subsequent tables relating to Soviet and Japanese accelerators. [23]

Table 7 presents the resultant data, giving a breakdown of world experimental high-energy physics publication output over the period 1969–80. [24] As can be seen, approximately 500 papers were published each year during this period (i.e. 1000 every two years). Of these, the three CERN accelerators were responsible for 26.5 percent (see the final column of the table). This is well over twice as many as the Brookhaven AGS accelerator (11.5 percent), three times the Fermilab figure (8.5 percent), nearly three-and-a-half times that of SLAC (8 percent), and some eight times the DESY figure (3 percent).

The next stage of our method of "converging partial indicators" is to evaluate the relative scientific productivity (i.e. output per unit of input) of the accelerators, and for this the figures on publication counts need to be set against those for inputs. The results of this analysis are shown in table 8. It can be seen that in the case of the CERN accelerators, their percentage of the total world output of experimental papers (26.5 percent) is very similar to their share of world resources (between a quarter and a third). The situation is similar for the DESY accelerators, whose share of world resources, though increasing from about 3 percent to 6 percent over the decade 1969–78 (see tables 2 and 4), has been matched by an equivalent growth in their share of the world's papers. The Brookhaven accelerator, in contrast, appears to have achieved a rather higher level of productivity – 11.5 percent of all publications compared with some 5–7 percent of funds and users. However, many of the papers that appeared in the early part of this decade were based on experiments carried out before 1969, so this comparison may in some respects be slightly misleading. Qualification is also needed in the case of Fermilab, for which, given its comparatively late entry into the field, the period 1973–78 provides the fairest indication of its productivity. In this period, its share of publication output (13.5 percent) is similar to its share of funds and users (11–13 percent). Finally, we can note that the

---

[18] See note c to table 7.

[19] See notes a and b to table 7 for the definitions used.

[20] See note e to table 7.

[21] Citations to preprints were, however, included in the citation analysis – see the next section for details of the way in which they were treated.

[22] In our previous work on electron high-energy physics (Martin and Irvine [18]), we relied primarily on publication lists provided by the research centres to derive data on numbers of publications. However, it has since become evident that the completeness of these lists varies across centres, largely because of the different procedures used by visiting scientists to report papers published in subsequent years after carrying out experiments. Their comprehensiveness also varies over time: the lists used for SLAC, and to a lesser extent DESY, for example, were found to be somewhat incomplete for the late 1960s and early 1970s.

[23] For further discussion of the problems of comparing the outputs of Soviet and Western accelerators, see Irvine and Martin [11].

[24] Although this paper is concerned primarily with the ten-year period from 1969 to 1978, some of the work carried out towards the end of this time was not published until 1979 or 1980. Figures for 1979–80 are therefore included.

Table 7
Numbers of experimental high-energy physics [a] papers [b] published in international journals [c] during the preceding two years

		1970	1972	1974	1976	1978	1980	Total 1969–78
Papers from W. European accelerators [e]	CERN	235 [d] (25%)	275 (27%)	265 (23.5%)	335 (28.5)	320 (29%)	235 25.5%)	1415 (26.5%)
	DESY	25 (2.5%)	35 (3.5%)	35 (3%)	30 (2.5%)	45 (4%)	60 (6.5%)	175 (3%)
	Others	80 (8.5%)	90 (9%)	105 (9%)	80 (6.5%)	65 (5.5%)	55 (6.5%)	420 (7.5%)
	Total W. Europe	340 (36%)	400 (39.5%)	410 (35.5%)	445 (38%)	435 (38.5%)	350 (38%)	2025 (37.5%)
Papers from US accelerators [e]	Brookhaven	180 (19%)	150 (14.5%)	155 (13.5%)	95 (8%)	60 (5.5%)	50 (5.5%)	635 (11.5%)
	Fermilab	–	5 (0.5%)	105 (9%)	175 (15%)	180 (16.5%)	180 19%)	470 8.5%)
	SLAC	55 (5.5%)	80 (8%)	80 (7%)	110 (9.5%)	105 (9.5%)	55 (6%)	425 (8%)
	Others	255 (27%)	210 (20.5%)	180 (15.5%)	105 (9%)	90 (8%)	45 (5%)	840 (15.5%)
	Total US	490 (51.5%)	445 (43.5%)	515 (45%)	485 (41%)	435 (39%)	330 (35.5%)	2370 (44%)
Papers from other accelerators [e] (Soviet Union, Japan)		120 (12.5%)	175 (17%)	225 (19.5%)	245 (21%)	250 (22.5%)	250 (27%)	1010 (18.5%)
World total		945 (100%)	1020 (100%)	1150 (100%)	1180 (100%)	1115 (100%)	930 (100%)	5410 (100%)

[a] High-energy physics is defined as "physics done with accelerators able to produce primary particles at an energy higher than 1 GeV" (Roche [31, p. 3]).

[b] We have attempted to use the same definition of an "experimental" high-energy physics paper as the international Particle Data Group – that is, the paper must "contain new (i.e. previously unpublished) experimental data .... If it is uncertain whether data is new, it is treated as new .... We exclude theoretical papers, unless a new particle or reaction property is derived from an analysis (of data) which is not based upon any particular phenomenological model .... Other subjects excluded are instrumentation, nuclear level structure, and studies of cosmic rays ...." (see Particle Data Group [28, p. 1.2]). However, unlike the Particle Data Group, we have excluded "reviews or compilations of data ..., preprints, book articles, conference proceedings, theses [and] experiment proposals" [28, p. 12].

[c] This publication list was derived by scanning the following 11 journals: *Lettere al Nuovo Cimento*, *Nuclear Physics B* (and before that, *Nuclear Physics*); *Nuovo Cimento*, *Physical Review D* (and before that, *Physical Review*); *Physical Review Letters*, *Physics Letters B* (and before that, *Physics Letters*); *Soviet Journal of Nuclear Physics*, *Soviet Physics – Doklady*; *Soviet Physics – JETP* and *JETP Letters*; and *Zeitschrift fur Physik C*. Additional information came from annual reports and publication lists produced by the various laboratories, and data compilations such as Particle Data Group [28] so the final publication list contains a small number of papers published in a variety of other journals.

[d] All totals have been rounded to the nearest 5, and all percentages to the nearest 0.5 percent.

[e] All the papers were scanned to establish which accelerator was used to obtain the experimental results. In a small number of cases (2.4 percent), more than one accelerator was used. For such cases, each accelerator used was credited with that paper, with the result that there is a small element of "double-counting" in the publication totals.

SLAC machines appear to have achieved a relatively high scientific productivity – 8.0 percent of all papers compared with 5–6 percent of funds and users.

However, as we have stressed earlier (see table 1), analyses based on publication counts alone can be misleading, since the publications from one accelerator may on average have had considerably more impact on the advance of knowledge than those from another. The next step, therefore, is to examine the relative overall impacts of those publications.

Table 8
Relative scientific productivity of major accelerators, 1969–78: Experimental publications in relation to inputs

	Inputs [a] (% of world total)	Experimental publications (% of world total)
CERN	~ 25–32 [b]	26.5
DESY	~ 4–5 [c]	3
Brookhaven	~ 5–7	11.5
Fermilab	~ 11–13 [b]	8.5 (13.5) [d]
SLAC	~ 5–6 [c]	8

[a] These figures have been taken from table 5, and represent an average of the figures for numbers of users and for funding.
[b] Except in the very early years of the decade, 1969–78.
[c] Except in the last year or so of the decade.
[d] Average figure for the period 1973–78.

6. Accelerator outputs – overall impact of scientific publications

Analysis of the relative frequency with which papers from one accelerator (or set of accelerators) are subsequently cited by scientists in other articles may be used to provide various indicators of their impact on the advance of scientific knowledge, though, as with other indicators, these are by no means unproblematic.[25] Three citation indicators – total numbers of citations, citations per unit input, and citations per paper – are of relevance here in assessing general scientific impact, while a fourth – numbers of highly cited papers – is used in the next section to evaluate more specifically which accelerators were responsible for the main breakthroughs in high-energy physics between 1969 and 1978.

The data-base from which all four citation indicators were constructed was compiled by manually scanning the *Science Citation Index* [33] for the years 1961–82. Using this method, full and relatively accurate citation records were obtained for the experimental publications from each accelerator. Manual scanning largely overcomes the technical problems of citation analysis listed in table 1 (mis-spelt names, incorrect volume or page numbers, etc.). And since the *Science Citation Index* covers the 11 main international high-energy physics journals, as well as most of the others occasionally used by experimentalists, the citation counts should (as was the case with the publication totals) be around 90–95 percent complete (again with the exception of the Soviet and East European figures[26]).

One special problem affecting citation analysis in high-energy physics should perhaps be mentioned. This is the extensive use of references to a preprint until the paper concerned is published.[27] (This practice ends virtually as soon as the paper is published, so it normally affects only the citations to a paper during its first year.) Such citations have all been credited to the paper that supersedes the preprint, and the same procedure has been adopted for references to articles "in press" and "to be published."

Comparative data on the overall level of citations gained by papers from the accelerators at each of the main Western high-energy physics centres are provided in table 9. Figures are given for even years only, and refer to the total numbers of citations made in those years to experimental journal articles published during the preceding four years,[28] i.e. the 1972 figures are for the number of citations in the 1972 edition of the *Science Citation Index*[29] to all papers published between 1969 and 1972. It can be seen that papers reporting results produced on CERN accelerators

---

[25] For a discussion of the intrinsic problems in citation analysis, see Martin and Irvine [19].

[26] See note a to table 9 and the discussion of this problem in Irvine and Martin [11].

[27] This problem is particularly pronounced for East European publications in which there is a far greater incidence of references to preprints than to published articles. This mode of referencing arises largely because of the frequently long delays before papers are published, but also because some East European research groups do not have easy access to journals and instead rely on informally circulated preprints. The Soviet citation totals must therefore be treated with *great caution* since they may represent an appreciable underestimate of the "true" citation totals (see Irvine and Martin [11]).

[28] Since most experimental high-energy physics papers have a peak citation-rate one to two years after publication, the use of a period longer than four years does not significantly affect the overall citation distribution, though it does tend to mask changes over time.

[29] While most of the citing articles in, say, the 1972 edition of the *Science Citation Index* will have been officially published in 1972, a few will be dated 1971 even though they actually appeared in 1972, too late for inclusion in the 1971 edition. Similarly, the 1973 edition will contain a few citing articles dated 1972. We have assumed that these two effects approximately cancel out when dealing with large numbers of publications.

Table 9
Numbers of citations[a] to experimental journal articles published during the preceding four years

		1972	1974	1976	1978	1980	Total 1972 + 1974 + 1976 + 1978
Citations to papers from W. European accelerators	CERN	1420[b] (27%)	2150 (31.5%)	1840 (24%)	2080 (25.5%)	1740 (28.5%)	7490 (26.5%)
	DESY	250 (4.5%)	180 (2.5%)	180 (2.5%)	450 (5.5%)	950 (15.5%)	1060 (4%)
	Others	420 (8%)	520 (7.5%)	440 (5.5%)	290 (3.5%)	240 (4%)	1680 (6%)
	Total W. Europe	2090 (39.5%)	2850 (42%)	2460 (32%)	2830 (34.5%)	2920 (48%)	10,230 (36.5%)
Citations to papers from US accelerators	Brookhaven	960 (18%)	760 (11%)	780 (10%)	420 (5%)	170 (3%)	2910 (10.5%)
	Fermilab	20 (0.5%)	780 (11.5%)	1900 (25%)	2620 (32%)	1320 (21.5%)	5320 (19%)
	SLAC	670 (12.5%)	670 (10%)	1330 (17%)	1250 (15%)	690 (11.5%)	3920 (14%)
	Others	1150 (21.5%)	990 (14.5%)	570 (7.5%)	570 (7%)	430 (7%)	3280 (11.5%)
	Total US	2800 (53%)	3190 (47%)	4590 (59.5%)	4850 (59.5%)	2620 (43%)	15,430 (55%)
Citations to papers from other accelerators (Soviet Union, Japan)		420 (8%)	740 (11%)	690 (9%)	510 (6%)	550 (9%)	2350 (8.5%)
World total		5310 (100%)	6780 (100%)	7740 (100%)	8190 (100%)	6090 (100%)	28,020 (100%)

[a] These figures have been derived using the *Science Citation Index*. This scans all 11 of the main international journals used in drawing up the publication list, as well as several of the other journals containing the occasional high-energy physics paper. Therefore, like the publication totals, the citation totals should be 90–95 percent complete, except in the case of papers from Soviet and Japanese accelerators. (Some Soviet papers are published in the physics journals of the various Soviet Republics, and these are not scanned by the *Science Citation Index*.)

[b] All totals have been rounded to the nearest 10, and all percentages to the nearest 0.5 percent.

earned 26.5 percent of all world experimental high-energy physics citations for the period 1969–78.[30] This is almost one-and-a-half times that of Fermilab's share (19 percent), twice the SLAC share (14 percent), two-and-a-half times Brookhaven's share (10.5 percent), and seven times that of DESY (4 percent). Thus, in terms of absolute numbers of citations, it would appear that the *total* impact of work carried out on CERN accelerators has been greater than that from any other laboratory over this period. Further comment will, however, be made on this question once the data on the distribution of highly cited papers have been taken into consideration.

It is important to recognize that variations in the total numbers of citations gained by the accelerators at different centres may reflect differences in activity levels as much as the relative significance of their respective experimental outputs. (We should remember, in particular, that comparison is being made between the outputs of one accelerator at Brookhaven and Fermilab, and those of three at CERN.) It is essential, therefore, to relate these citation totals to the inputs for the various accelerators. The relevant figures are given in table 10 and suggest that, in terms of citations per unit input, the SLAC accelerators have had the best record, followed by the Fermilab and Brookhaven accelerators, with the CERN and DESY machines some way behind.

Probably the most useful of the indicators based on total citations is the average number of citations per paper. This again takes into account

---

[30] As noted above, only citations in even years are reported here. However, the inclusion of citations made in the intervening odd years does not lead to appreciably different results.

Table 10
Relative scientific impact of major accelerators, 1969–78: Citations in relation to inputs

	Inputs [a] (% of world total)	Citations to recent papers (% of world total)
CERN	25–32 [b]	26.5
DESY	4–5 [c]	4
Brookhaven	5–7	10.5
Fermilab	11–13 [b]	19 (23.5) [d]
SLAC	5–6 [c]	14

[a] These figures have been taken from table 5, and represent an average of the figures for numbers of users and for funding.
[b] Except in the very early years of the decade, 1969–78.
[c] Except in the last year or so of the decade.
[d] Average of the totals for 1974, 1976 and 1978.

differences in the scale of resources and research activity associated with each accelerator, and enables direct comparisons to be made of the average impacts of the great mass of experimental papers that at best contribute only marginally to scientific progress. Table 11 gives the relevant data for various four-year periods between 1969 and 1980; the 1972 figures, for example, represent the number of citations in the 1972 *Science Citation Index* to experimental papers published between 1969–72, divided by the total number of those papers. In terms of this indicator, the CERN accelerators again come out behind those at SLAC and Fermilab,[31] as well as behind those at DESY, but ahead overall of the Brookhaven AGS – a relatively old accelerator yielding publications with a somewhat lower than average impact in the latter part of this period.

7. Accelerator outputs – highly cited papers and discoveries

We pointed out previously that, while indicators based on publication counts and total or average numbers of citations provide information on the vast number of small incremental contributions to scientific progress, they may reveal very little indeed about which accelerators have been responsible for the occasional crucial discoveries that have completely transformed high-energy physics. In order to focus on such discoveries (and also on slightly lower-level, but nevertheless important, advances), it is useful to examine the data on highly cited papers.[32]

First, however, the period 1969–78 needs to be set in its historical context. While several major discoveries in high-energy physics had been made in the second half of the 1950s and the early 1960s, between 1965 and 1968 there were virtually none, and from 1969 to 1972 only a few (deep-inelastic scattering at SLAC,[33] the first indications of rising total cross-section in hadron collisions at Serpukhov, and some early results from the ISR at CERN).

In contrast, the period 1973–77 resulted in numerous major discoveries – most notably those of neutral currents, the J/psi particle, the heavy lepton tau, charmed particles, and the upsilon – which completely transformed the nature of particle physics.[34] High-energy physics during this period, and in particular at the time of the so-called "November revolution" in 1974 (when the J/psi was discovered), may be seen as an epoch of truly "revolutionary" science while the previous period from 1965–72 was more one of 'normal' science (cf. Kuhn [16]).

During this revolutionary period, certain key experimental papers clearly achieved an impact many times greater than that of the great majority of other publications. It is therefore necessary to look at the distribution of highly cited papers between accelerators, since this may provide a more relevant indicator of relative contributions to scientific knowledge than total numbers of publications or citations. These data are given in table 12.

Over the decade 1969–78, approximately 5400

---

[31] The high figure for the Fermilab accelerator is partly due to several controversial papers subsequently found to be "mistaken" – see the following section on highly cited papers.

[32] The only high-energy physics experiment during the 1960s to have subsequently been awarded a Nobel Prize was that which discovered CP-violation. The paper reporting this discovery was the most-cited experimental publication of that decade. Similarly, the most-cited experimental papers of the 1970s concern the discovery of the J/psi particle, and that too was rewarded by a Nobel Prize.

[33] This work was published in 1969, although the experiment had been carried out the previous year. In what follows, the dates given for experiments refer to when the results were published rather than when the experiment was run.

[34] The results from these and other experiments during the 1960s and 1970s are discussed in more detail in paper II [12].

BEST COPY AVAILABLE

Table 11
Average citations per paper (CPP) for journal articles published during the preceding four years

		1972	1974	1976	1978	1980	Average for 1972, 1974, 1976 and 1978
CPP for papers from W. European accelerators	CERN	2.8	4.0	3.0	3.2	3.1	3.2
	DESY	4.0	2.5	2.7	5.7	8.8	3.8
	Others	2.5	2.6	2.4	2.0	2.0	2.4
	W. European average CPP	2.8	3.5	2.9	3.2	3.7	3.1
CPP for papers from US accelerators	Brookhaven	2.9	2.5	3.2	2.7	1.6	2.8
	Fermilab	(3.8) [a]	6.9	6.7	7.3	3.7	7.0
	SLAC	4.9	4.2	7.1	5.8	4.4	5.6
	Others	2.5	2.5	2.0	2.9	3.2	2.5
	US average CPP	3.0	3.3	4.6	5.3	3.4	4.0
CPP for papers from other accelerators (Soviet Union, Japan)		1.4	1.9	1.5	1.0	1.1	1.4
World average CPP		2.7	3.1	3.3	3.5	3.0	3.2

[a] This value is based on only five papers, and may not therefore be very significant.

experimental journal articles were published. Of these, the top 0.2 percent most cited earned 100 or more citations ($n \geq 100$) in any one year. Examination of these 11 papers (see the final column of table 12) reveals that they correspond closely with the crucial discoveries listed above.[35] Furthermore, it can be seen that the SLAC accelerators produced over half of these papers (6 out of 11), many times more than the accelerators at any other centre; the CERN, Brookhaven, Fermilab and Serpukhov machines achieved only one each. Such a striking difference would suggest that, in this period of revolutionary science, the SLAC accelerators were central in advancing our knowledge of high-energy physics.[36]

It is also instructive to examine the distribution between accelerators of discoveries of a slightly lesser magnitude (the term "major advances" will be used, in distinction from "crucial discoveries") by considering the data on papers cited more than 50 and more than 30 times in a year (see the middle two columns of table 12). These correspond, respectively, to the top 0.8 percent and top 2 percent most-cited papers over the decade 1969–78. As can be seen, the Fermilab accelerator has yielded most papers at these levels, followed by the CERN and SLAC accelerators. However, it should be noted that at least four of the Fermilab papers (two on "trimuons" and two on the "high-y anomaly") earning 50 or more citations are now considered by most high-energy physicists to have been "mistaken," and to have seriously misled theorists for several years before they were refuted by more accurate results from the CERN SPS accelerator. At the time these papers appeared, high-energy physicists were beginning to regard the "Weinberg–Salam theory" as a strong candidate for a unified theory of electromagnetic and weak interactions. Yet the Fermilab results appeared to imply that the simple model of Weinberg–Salam was invalid and a more complicated version was required. For a short period thereafter, these papers, not surprisingly, had a rather high impact within the scientific community (reflected in their citation rates), as the following review article of the time makes clear:

> Perhaps *the most dramatic development* in neutrino physics in the last year or so has been the observation of substantial anomalies in the total and differential cross-sections at high-energy. … The new data … appear to lead *inexorably* to new quarks and new couplings, simply because other mechanisms are inadequate (Perkins [29, p. 470 and p. 475], emphasis added).

Understandably, some of those who laboured long

---

[35] The degree of correspondence is examined in detail in paper II [12]. However, it is worth stressing here that *none* of the 11 is a "mistaken" paper.

[36] The same conclusion was reached in Martin and Irvine [18], but in a much more tentative form.

Table 12
Numbers [a] of highly cited papers (HCP) for the period 1969–78: Numbers of papers cited n or more times in one year

		$n \geqslant 15$	$n \geqslant 30$	$n \geqslant 50$	$n \geqslant 100$
HCP for W. European accelerators	CERN	111 (26%) [b]	31 (26%)	9 (19.5%)	1 (9%)
	DESY	20 (4.5%)	9 (7.5%)	3 (6.5%)	0 (0%)
	Others	14 (3.5%)	4 (3.5%)	1 (2%)	1 (9%)
	Total W. Europe	145 (34%)	44 (37%)	13 (28%)	2 (18%)
HCP for US accelerators	Brookhaven	37 (8.5%)	6 (5%)	2 (4.5%)	1 (9%)
	Fermilab	106 (24.5%)	37 (31%)	17 (37%)	1 (9%)
	SLAC	75 (17.5%)	21 (17.5%)	11 (24%)	6 (54.5%)
	Others	42 (10%)	7 (6%)	1 (2%)	0 (0%)
	Total US	260 (60.5%)	71 (59.5%)	31 (67.5%)	8 (72.5%)
HCP for other accelerators		24 (5.5%)	4 (3.5%)	2 (4.5%	1 (9%)
World total		429 (100%)	119 (100%)	46 (100%)	11 (100%)

[a] These figures have been obtained by scanning the *Science Citation Index* for the years 1969 to 1982. The likely errors should be rather smaller than the 10 percent figure of previous tables.

[b] All percentages have been rounded to the nearest 0.5 percent.

and hard in constructing new theoretical models to accommodate the errant Fermilab data are now somewhat resentful that their efforts appear to have been needless.[17] One eminent theorist subsequently described the time between the first appearance of the anomalous results and their eventual refutation as

> a confusing period of exhilaration and disappointment, alarms and excursions. Experiment confirmed the [Weinberg–Salam] theory; experiment denied the theory. Enormous theoretical effort was devoted to producing grotesque mutant versions of the theory consistent with the new experimental results; the new experiments were shown to be in error; the mutants were slain. (This happened at least three times: with high-y anomalies, with trimuons, and with atomic parity violation). In the last few years, though, the experimental situation seems to have stabilized in agreement with the original 1971 version of the theory. The Weinberg–Salam model is now the standard theory of the electro-weak interaction (Coleman [6, p. 122]).

It is far from simple to decide how such "mistakes" should be assessed. Certainly, they did initially have a major impact on high-energy physics and their citation figures for the period reflect this. However, as soon as they were generally recognized to be "mistaken," their impact fell dramatically to become almost negligible. (This was reflected in their citation rates rapidly dropping to virtually zero.) With hindsight, that initial impact might now be judged to have been negative (i.e. retarding rather than advancing scientific knowledge). According to another interpretation, however, the efforts to explain these results were not in vain: thus, although the Fermilab data led theorists to posit the existence of a new, fifth type of quark for reasons that turned out to be specious, they

> nonetheless served a valuable heuristic purpose in stimulating speculation about the properties of particles composed of heavier quarks, properties the upsilon [an important new

---

[17] This came across very strongly in several of the interviews that we conducted. One distinguished theorist commented: "The high-y anomaly attracted a tremendous amount of theoretical activity which was just misguided. So they did a great disservice to the high-energy physics community. They held things back by several years" (Interview, 1981).

AVAILABLE

> particle discovered a little later] turned out to have (Lederman [17, p. 67]).

Under this latter interpretation, these "mistaken" papers clearly did contribute to the advance of scientific knowledge, although probably not by as much as their subsequent citation records would suggest. The Fermilab data on highly cited papers in table 12 and subsequent tables should therefore be viewed in this light.

We can go on from here to consider the distribution across accelerators of a still lower level of discoveries (the term "advances" will be used). The first column of table 12 presents data on those papers earning 15 or more citations in a year (corresponding to the top 8 percent most-cited papers). In marked contrast with the data on "crucial discoveries" and "major advances," it can be seen that the CERN accelerators produced 111 such papers, slightly more than the Fermilab accelerator (106) and significantly more than the SLAC (75), Brookhaven (37) and DESY (20) machines.

Finally, we should examine data relating to the different "lifetimes" of highly cited papers. This is a necessary dimension to any assessment of impact since it is clear that some key publications have a continuing major impact over several years, while others (most notably "mistaken" papers) are relatively short-lived. This effect can be allowed for by analyzing the number of times that highly cited papers earn more than a certain number of citations in a year. The relevant data are given in table 13. Overall, we can conclude that the pattern of distribution between the accelerators is not markedly different from that evident in table 12, although the figures for the Fermilab accelerator for papers cited 50 or more times are reduced from 37 percent to 30 percent of the respective world totals. This is largely because of the shorter lifetimes of the mistaken Fermilab papers.

How do these various data on highly cited papers relate to the inputs for the various accelerators? Table 14 presents data on relative scientific impact, setting the figures for $n \geq 15$ in table 13 against the relevant input figures. (Comparisons of inputs with data on more highly cited papers in

Table 13
Numbers of highly cited papers (HCP) for the period 1969-78: Number of times highly cited papers received n or more citations in a year [a]

		$n \geq 15$	$n \geq 30$	$n \geq 50$	$n \geq 100$
HCP for W. European accelerators	CERN	269 (27.5%) [b]	56 (23.5%)	16 (18.5%)	1 (5.5%)
	DESY	43 (4.5%)	12 (5%)	3 (3.5%)	0 (0%)
	Others	32 (3.5%)	6 (2.5%)	3 (3.5%)	2 (11%)
	Total W. Europe	344 (35%)	74 (31%)	22 (25.5%)	3 (16.5%)
HCP for US accelerators	Brookhaven	83 (8.5%)	18 (7.5%)	6 (7%)	3 (16.5%)
	Fermilab	232 (23.5%)	71 (30%)	26 (30%)	2 (11%)
	SLAC	187 (19%)	57 (24%)	27 (31%)	9 (50%)
	Others	79 (8%)	8 (3.5%)	2 (2.5%)	0 (0%)
	Total US	581 (59%)	154 (64.5%)	61 (70%)	14 (78%)
HCP for other accelerators (Soviet Union, Japan)		59 (6%)	10 (4%)	4 (4.5%)	1 (5.5%)
World total		984 (100%)	238 (100%)	87 (100%)	18 (100%)

[a] These figures have been obtained by scanning the *Science Citation Index* for the years 1969 to 1982.

[b] All percentages have been rounded to the nearest 0.5 percent.

Table 14
Relative scientific impact of major accelerators, 1969–78
Highly cited papers in relation to inputs

	Inputs [a] (% of world total)	Number of times papers earned ≥ 15 citations in a year (% of world total)
CERN	$\sim$ 25–32 [b]	27.5
DESY	$\sim$ 4–5 [c]	4.5
Brookhaven	$\sim$ 5–7	8.5
Fermilab	$\sim$ 11–13 [b]	23.5
SLAC	$\sim$ 5–6 [c]	19

[a] These figures have been taken from table 5, and represent an average of the figures for numbers of users and for funding.

[b] Except in the very early years of the decade, 1969–78.

[c] Except in the last year or so of the decade.

both tables 12 and 13 could also, of course, be made.) Again, the SLAC accelerators emerge best from this comparison of input-adjusted impact, followed by the Fermilab (but with the reservation about mistaken papers) and Brookhaven machines.

The output from the CERN accelerators in terms of this indicator (at $n \geq 15$) is little different from their share of inputs, a conclusion that needs some qualification if data on very highly cited papers ($n \geq 100$) are introduced into the comparison.

8. Accelerator outputs – peer-evaluation

Our previous studies of various Big Science specialities have strongly suggested that, while indicators based on publication and citation analysis provide essential information on the relative contributions to scientific knowledge associated with different research facilities, such indicators always need to be considered alongside peer-evaluation data. So how do high-energy physicists judge the magnitude of the contributions from the CERN accelerators compared with those from other accelerators?

Peer-evaluation data were obtained for all the accelerators included in the present study through interviews with 182 [38] experimental and theoretical high-energy physicists. These were carried out in the latter part of 1981 and the first half of 1982. In selecting our interview sample, care was taken to ensure that as far as was possible the views of researchers from Western Europe, Eastern Europe and the United States were all well represented. [39] The result was that 41 high-energy physicists were interviewed at CERN, 71 in CERN user-groups in Western Europe, 28 at Brookhaven (and in the user-group at the State University of New York, Stonybrook), 24 at Fermilab, and a total of 18 in Bulgaria, Finland and Poland.

The interviews were intensive (typically lasting 1½ to 2 hours) and structured, being based on a common set of questions but with a few additional questions for special groups of researchers. After giving brief details of their background and career, interviewees were asked to describe their own research work and perceived contributions to high-energy physics. The questions were then gradually broadened in several stages, interviewees being requested first of all to identify the principal contributions of the various collaborations in which they had worked, then those from the accelerators they had used, and, finally, the overall contributions from the world's other major accelerator facilities. In some cases, memories had to be jolted or occasionally corrected when discoveries or major research programmes were attributed to the wrong accelerators. [40] Apart from its information value, this exercise in systematically recollecting the contributions of different accelerators was useful in helping respondents prepare themselves for the peer-review section of the interview. In this, interviewees were invited to rank the accelerators at fourteen different laboratories according to the relative magnitude of their contributions to high-energy physics over the period 1969–78. (For this question, all three CERN accelerators were treated as one unit, [41] as was the case with other multi-accelerator laboratories.) In addition to the five major Western laboratories already discussed, we included in the peer-ranking three smaller facilities in the United States (Argonne; the Cambridge

---

[38] This does not include a number of other interviews that were terminated prematurely when it became apparent that interviewees had insufficient knowledge to answer the questions satisfactorily, being unable, for example, to recollect correctly which accelerators were responsible for certain experimental results.

[39] See note 45, below.

[40] For example, some physicists wrongly attributed the discovery of neutral currents to the CERN SPS rather than the PS.

[41] However, it should be re-emphasized that we are here assessing the outputs from the CERN *accelerators*, not the performance of CERN as a centre nor its staff.

Electron Accelerator, CEA; and Cornell), two in Britain (the Nimrod and NINA accelerators at Rutherford and Daresbury Laboratories respectively), and four in the Soviet Union (the large accelerator at Serpukhov, and the smaller machines at Dubna, Moscow, and Yerevan). Thus all proton accelerators with an energy of 7 GeV or greater, and all electron accelerators of 4 GeV or more, operating between 1969 and 1978,[42] were included.[43]

Of the 182 physicists interviewed, relatively few were able to rank in full order from 1 (top) to 14 (bottom) the accelerators at all fourteen laboratories. Most regarded at least two or more as having made equivalent contributions, particularly in the case of those at the lower end of the order. Others preferred to identify five or six distinct groups of accelerators and then to rank these in order. Only 8 percent found it too difficult to undertake the ranking, or refused for personal or professional reasons.

Average rankings on a scale of 1 to 14 were then calculated,[44] and the results are given in table 15.[45] In addition to presenting "overall average rankings" for each research facility (see the penultimate column of the table), we have classified the judgements of the 168 high-energy physicists who provided rankings into six groups in order to establish whether interviewees' current institutional affiliation had any significant effect upon the results. These groups consist of those interviewed in (1) Brookhaven and Stonybrook; (2) Fermilab; (3) CERN; (4) research teams in three large CERN Member States (Britain, France and Italy); (5) research teams in two small CERN Member States (the Netherlands and Norway); and (6) user-groups of Soviet accelerators in Bulgaria, Finland and Poland.

Certain comments can be made about the figures in table 15. First, a high degree of consistency exists between the average rankings for each of the six groups; with just two exceptions, all are within one unit of the overall average (for all 168 interviewees), and over three-quarters are within half a unit. Second, those two exceptions are the rankings by Finnish and East European physicists of the scientific contributions associated with the Dubna and Serpukhov accelerators in the Soviet Union. Given the problems of scientific communication between East and West, this larger difference is probably due to a certain ignorance on the part of many scientists in the West of the experimental work of these accelerators.[46] Third, a small "self-ranking" effect seems to be evident in certain cases; for example, high-energy physicists interviewed at Brookhaven ranked the AGS accelerator 3.1 compared with an overall average of 3.7, while the equivalent figures for Fermilab were 2.3 and 3.1. Table 16 attempts to set out in a more explicit way the effects of self-ranking (see the first two columns); thus, the 12 interviewees who, for example, had used the Argonne accelerator ranked it 7.5 ("self-ranking") while the remaining 156 (who had not) ranked it 7.6 ("peer-ranking"). However, the effect is not, in general, sufficiently large to cast doubt on the reliability of the results this approach yields. Fourth, and finally, although a number of senior physicists interviewed felt that certain categories of physicists – in particular, older, established researchers (i.e.

---

[42] Accelerators that only began to yield published results after 1978 (such as PETRA at DESY, and CESR at Cornell) have been excluded.

[43] Conversely, accelerators with less than these energies were excluded. It should be noted, however, that several of them, for example those at Berkeley, Orsay, Saclay, Frascati, and Novosibirsk, might well have been placed above some of those ranked towards the bottom of our list of fourteen had they been included.

[44] Where, for example, two accelerators were ranked first equal, they were each given the ranking 1.5 (i.e. (1 + 2)/2); where three were placed first equal, they were ranked 2 (i.e. (1 + 2 + 3)/3); and so on.

[45] It should be stressed that these rankings were made primarily by physicists whose main research experience has been with proton accelerators. This was necessary given that the aim of the study was to evaluate the scientific outputs of the CERN accelerators in relation to their main competitors. Similarly, in our previous study of world electron accelerators, those chosen to undertake the peer-ranking exercise were mainly electron physicists. (See table XVII of Martin and Irvine [18] for the relevant results.) However, it should be noted that the results are strikingly similar, despite the different cognitive and institutional locations of the scientists making the peer-judgements. In the electron-accelerator study, only thirteen facilities were ranked (Cornell was not included), so some slight adjustment needs to be made to the rankings in order to compare the results directly.

[46] The difference between the East European rankings of the Dubna and Serpukhov accelerators and the "overall average" rankings is much bigger than the "self-ranking" effects (see text above) in the results for Western accelerators. This suggests that at least part of the difference must be attributed to an "ignorance" effect. See Irvine and Martin [11] for further discussion.

Table 15
Peer-evaluation rankings[a] of the contributions to high-energy physics[b] made by accelerators at 14 laboratories between 1969 and 1978

Accelerator(s) (energy)	Peer-evaluation rankings[c] by high-energy physicists in						Overall average rankings	Relative position
	Brookhaven and Stonybrook	Fermilab	CERN	Large CERN Member States	Small CERN Member States	Bulgaria, Finland Poland		
(Sample size)	(26)	(23)	(39)	(45)	(21)	(14)	(168)	
Proton machines								
Argonne ZGS	6.9	6.6	8.5	7.6	7.6	7.8	7.6	6
(12 GeV)	(±0.3)	(±0.2)	(±0.3)	(±0.3)	(±0.4)	(±0.5)	(±0.1)	
Brookhaven AGS	3.1	4.1	3.9	3.7	3.4	3.6	3.7	4
(33 GeV)	(±0.2)	(±0.2)	(±0.2)	(±0.2)	(±0.3)	(±0.3)	(±0.1)	
CERN (i) PS (28 GeV)								
(ii) ISR(31 + 31 GeV)	2.5	3.7	2.3	2.4	2.3	2.0	2.4	2
(iii) SPS (400 GeV)	(±0.1)	(±0.1)	(±0.1)	(±0.1)	(±0.2)	(±0.2)	±0.1)	
Dubna	11.9	12.2	11.8	11.4	11.0	10.3	11.5	12
(10 GeV)	(±0.3)	(±0.2)	(±0.2)	(±0.2)	(±0.3)	(±0.5)	(±0.1)	
Fermilab	3.3	2.3	3.1	3.4	3.6	3.0	3.1	3
(400 GeV)	(±0.1)	(±0.2)	(±0.1)	(±0.1)	(±0.2)	(±0.3)	(±0.1)	
Moscow ITEP	12.1	13.0	12.2	12.1	11.8	12.2	12.2	13=
(7 GeV)	(±0.3)	(±0.1)	(±0.2)	(±0.2)	(±0.3)	(±0.3)	(±0.1)	
Rutherford Nimrod	9.1	9.4	8.9	8.6	9.9	10.1	9.1	9=[d]
(7 GeV)	(±0.4)	(±0.3)	(±0.2)	(±0.3)	(±0.3)	(±0.5)	(±0.1)	
Serpukhov	9.4	8.6	9.1	8.9	7.8	7.1	8.7	7=[d]
(70 GeV)	(±0.4)	(±0.3)	(±0.3)	(±0.3)	(±0.5)	(±0.6)	(±0.2)	
Electron machines								
CEA (i) 6 GeV	9.0	8.7	9.0	9.1	10.1	10.2	9.2	9=[d]
(ii) BYPASS (3 + 3 GeV)	(±0.3)	(±0.5)	(±0.3)	(±0.3)	(±0.3)	(±0.4)	(±0.1)	
Cornell	8.4	7.9	8.1	8.9	8.6	9.1	8.5	7=[d]
(12 GeV)	(±0.3)	(±0.3)	(±0.3)	(±0.3)	(±0.5)	(±0.5)	(±0.1)	
Daresbury NINA	9.6	10.7	9.5	9.7	10.5	10.7	10.0	11
(4 GeV)	(±0.3)	(±0.3)	(±0.3)	(±0.2)	(±0.3)	(±0.5)	(±0.1)	
DESY (i) (7 GeV)	6.3	5.0	5.0	5.7	4.9	5.4	5.4	5
(ii) DORIS (5 + 5 GeV)	(±0.4)	(±0.2)	(±0.2)	(±0.3)	(±0.3)	(±0.3)	(±0.1)	
SLAC (i) (20 GeV)	1.3	1.3	1.3	1.1	1.4	1.6	1.3	1
(ii) SPEAR (4 + 4 GeV)	(±0.1)	(±0.1)	(±0.1)	(±0.1)	(±0.1)	(±0.2)	(±0.1)	
Yerevan	12.1	12.5	12.3	12.3	12.1	11.9	12.2	13=
(6 GeV)	(±0.2)	(±0.3)	(±0.2)	(±0.2)	(±0.3)	(±0.4)	(±0.1)	

[a] 1 = highest ranking; 14 = lowest ranking.

[b] Contributions to other areas of research such as nuclear physics or synchrotron radiation are *not* being assessed here.

[c] The figure in brackets indicate the root-mean-square variations between the assessments made by the high-energy physicists, giving some approximate idea of the divergence of opinion within the different groups.

[d] Differences of only 0.1 or 0.2 in the overall average rankings are not statistically significant.

those who have been active for 15 or 20 years), theorists (who are generally less "committed" to particular accelerator centres), and experimentalists who had used a large number of accelerators – would be likely to give more balanced and reliable rankings, we found little evidence that this was the case. In this respect, table 16 (final two columns) considers the rankings made by those gaining their Ph.D. degrees before and after 1965, while table 17 presents data on the judgements made by theorists as well as figures "weighting" individual physicists' rankings by the number of accelerators used. As can be seen, these groups of researchers hold relatively similar views on accelerator performance to those of high-energy physicists in general.

Table 16
Rankings [a] of accelerators for 1969 to 1978: Effects of self-ranking and age [b].

Accelerator(s) ([illegible])	Self-ranking	Peer-ranking	Ph.D. awarded before (or during) 1965 ($n = 89$)	Ph.D. after 1965 (or no Ph.D.) ($n = 79$)
Proton machines				
Argonne ZGS (12 GeV)	7.5(±0.5) ($n = 12$)	7.6(±0.1) ($n = 156$)	7.7(±0.2)	7.5(±0.2)
Brookhaven AGS (33 GeV)	3.5(±0.2) ($n = 47$)	3.7(±0.1) ($n = 121$)	3.8(±0.1)	3.5(±0.1)
CERN (i) PS (28 GeV) (ii) ISR (31 + 31 GeV) (iii) SPS (400 GeV)	2.3(±0.1)	2.6(±0.1)	2.3(±0.1) ($n = 108$)	2.6(±0.1) ($n = 60$)
Dubna (10 GeV)	11.2(±0.4) ($n = 9$)	11.5(±0.1) ($n = 159$)	11.6(±0.2)	11.4(±0.2)
Fermilab (400 GeV)	2.6(±0.1) ($n = 45$)	3.3(±0.1) ($n = 123$)	3.1(±0.1)	3.2(±0.1)
Moscow ITEP (7 GeV)	–	12.2(±0.1) ($n = 168$)	12.2(±0.1)	12.2(±0.1)
Rutherford Nimrod (7 GeV)	7.5(±0.4) ($n = 23$)	9.4(±0.1) ($n = 145$)	8.8(±0.2)	9.5(±0.2)
Serpukhov (70 GeV)	6.7(±0.4) ($n = 19$)	9.0(±0.2) ($n = 149$)	8.7(±0.2)	8.7(±0.2)
Electron machines				
CEA (i) 6 GeV (ii) BYPASS (3 + 3 GeV)	–	9.2(±0.1) ($n = 165$)	9.3(±0.2)	9.1(±0.2)
Cornell (12 GeV)	–	8.5(±0.1) ($n = 165$)	8.4(±0.2)	8.5(±0.2)
Daresbury NINA (4 GeV)	–	10.0(±0.1) ($n = 163$)	9.9(±0.2)	10.0(±0.2)
DESY (i) 7 GeV (ii) DORIS (5 + 5 GeV)	5.1(±0.4) ($n = 17$)	5.4(±0.1) ($n = 151$)	5.5(±0.2)	5.3(±0.2)
SLAC (i) 20 GeV (ii) SPEAR (4 + 4 GeV)	1.1(±0.1) ($n = 17$)	1.3(±0.1) ($n = 151$)	1.3(±0.1)	1.2(±0.1)
Yerevan (6 GeV)	–	12.2(±0.1) ($n = 168$)	12.2(±0.1)	12.3(±0.1)

[a] 1 = highest ranking, 14 = lowest ranking. The figures in brackets indicate the root-mean-square variations between the assessments made by the high-energy physicists, giving some approximate idea of the divergence of opinion within the different groups.

[b] The sample size of each group is denoted by n.

One of the main difficulties with peer-evaluation is, of course, the problem of overcoming researchers' natural worries about the implications for themselves and colleagues of negative judgements, particularly in a period of financial pressure on funding bodies.[47] However, given the comparatively small systematic variations found in the rankings made by the very different groups of high-energy physicists, it would seem fair to conclude that this problem has largely been overcome by our peer-evaluation technique, and that a certain degree of confidence can be placed in the results. This said, we are now in a position to summarize the peer-review findings. The final column of table 15 converts the overall average rankings made by physicists back into the relative positions of the accelerators at the fourteen laboratories. The clear conclusion here is that the SLAC accelerators are regarded as having contributed most to the advance of high-energy physics over the period 1969–78 (see note a to table 15), followed by the accelerators at CERN, Fermilab,

---

[47] See Irvine and Martin [13] for a discussion of the problems that now confront peer-review as a mechanism for scientific management.

Table 17
Rankings of accelerators [a] for 1969 to 1978: Views of theorists and rankings weighted by number of accelerators used [b]

Accelerator(s) (energy)	Theorists ($n = 21$)	Rankings weighted by number of accelerators used	Overall unweighted average rankings ($n = 168$)
Proton machines			
Argonne ZGS (12 GeV)	8.2(±0.4)	7.6(±0.1)	7.6(±0.1)
Brookhaven AGS (33 GeV)	3.3(±0.2)	3.8(±0.1)	3.7(±0.1)
CERN (i) PS (28 GeV) (ii) ISR (31 + 31 GeV) (iii) SPS (400 GeV)	2.3(±0.1)	2.4 [c]	2.4(±0.1)
Dubna (10 GeV)	11.[illegible](±0.4)	11.6(±0.1)	11.5(±0.1)
Fermilab (400 GeV)	3.5(±0.2)	3.1 [c]	3.1(±0.1)
Moscow ITEP (7 GeV)	11.9(±0.3)	12.4(±0.1)	12.2(±0.1)
Rutherford Nimrod (7 GeV)	[illegible].8(±0.4)	8.9(±0.1)	9.1(±0.1)
Serpukhov (70 GeV)	8.[illegible](±0.5)	8.7(±0.1)	8.7(±0.2)
Electron machines			
CEA (i) 6 GeV (ii) BYPASS (3 + 3 GeV)	[illegible]	9.3(±0.1)	9.2(±0.1)
Cornell (12 GeV)	8.9(±0.4)	8.5(±0.1)	8.5(±0.1)
Daresbury NINA (4 GeV)	10.1(±0.3)	9.9(±0.1)	10.0(±0.1)
DESY (i) 7 GeV (ii) DORIS (5 + 5 GeV)	5.5(±0.4)	5.3(±0.1)	5.4(±0.1)
SLAC (i) 20 GeV (ii) SPEAR (4 + 4 GeV)	1.2(±0.1)	1.3 [c]	1.3(±0.1)
Yerevan (6 GeV)	11.9(±0.3)	12.4(±0.1)	12.2(±0.1)

[a] 1 = highest ranking, 14 = lowest ranking. The figures in brackets indicate the root-mean-square variations between the assessments made by the high-energy physicists, giving some approximate idea of the divergence of opinion within the different groups.

[b] The sample size of each group is denoted by n.

[c] Root-mean-square variation of less than 0.05.

Brookhaven and DESY, in that order. As stated earlier (note 45), these findings are closely in line with the opinions of electron high-energy physicists reported in Martin and Irvine [18].

9. Accelerator outputs – an overall assessment of the period 1969–78

In concluding our assessment, it is necessary to consider the overall picture of the relative scientific performance between 1969 and 1978 of the CERN accelerators and their users that is provided by the different partial indicators taken together. Table 18 summarizes the results for the five main Western laboratories. The CERN accelerators, for example, have been responsible for the greatest number of publications and citations, and, in terms of these two indicators, their relative position is therefore first. However, in terms of citations per paper the CERN accelerators come behind the Fermilab, SLAC, and DESY machines,

BEST COPY AVAILABLE

in fourth position for the ten years under consideration.

The peer-evaluation results represent the views of scientists who, in according relative positions to laboratories, have tried to balance small numbers of crucial discoveries (for which the number of papers cited 100 or more times in a year provides a useful indicator) against overwhelmingly greater numbers of lower-level contributions (for which citation totals provide a reasonable indication of overall impact).[48] Thus, it is notable that, while the CERN accelerators rank highest in terms of both total citations and "advances" (papers cited 15 or more times in a year), the physicists interviewed nevertheless ranked them second for the decade in question, behind the SLAC machines. This suggests that, in considering a period which witnessed revolutionary change in their subject, they attached greater weight to the criterion of whether or not accelerators have been responsible for the "crucial discoveries" that transformed the field. In line with the figures on very highly-cited papers ($n \geqslant 100$) showing SLAC as having been responsible for over half such crucial discoveries, they ranked the Stanford accelerators ahead of those at CERN. In addition, among the five facilities, only DESY apparently failed to produce a single such crucial discovery during the period concerned,[49] and those interviewed ranked it fifth. The CERN, Brookhaven and Fermilab facilities cannot be distinguished in terms of crucial discoveries (between 1969 and 1978, each yielded one paper cited more than 100 times); however, the indicators of lower-level contributions (based on total citations and "advances") both suggest that Fermilab should be ranked ahead of Brookhaven, but behind CERN, agreeing with the relative positions accorded to these three by the peer-reviewers.

In the light of these results, certain comments should perhaps be made about the notion of "convergence" between partial indicators that was introduced in our earlier work on assessing large basic research facilities (particularly in Martin and Irvine [19]). It is clear that, in the present study, the indicators do not all converge, and the set of indicators used needs to be interpreted with some care. In part, this lack of overall convergence is due to the fact that the various facilities being compared differ rather more in terms of their relative sizes and inputs than those examined previously.[50] But probably more important is the fact that, in a field of science undergoing a period of revolutionary change, as did high-energy physics during the mid-1970s, indicators based on publication and citation totals lose much of their utility since they tend not to reveal where the crucial discoveries were made. Citations per paper can also be misleading unless great care is taken to identify, and allow for, the rather greater numbers of "mistaken" papers that appear to surface during such periods of theoretical uncertainty. Even

Table 18

Accelerator output, 1969–78: Summary of the relative positions[a] suggested by the various partial indicators

	Publications	Total citations	Citations per paper	"Advances"[b]	"Crucial discoveries"[c]	Peer-evaluation
CERN	1	1	4	1	3 =	2
DESY	5	5	3	5	5	5
Brookhaven	2	4	5	4	3 =	4
Fermilab	3	2	1	2	3 =	3
SLAC	4	3	2	3	1	1

[a] 1 = highest ranking, 5 = lowest ranking.

[b] Based on the number of papers cited 15 or more times in a year (the top 8 percent most-cited papers).

[c] Based on the number of papers cited 100 or more times in a year (the 0.2 percent most-cited papers).

---

[48] In paper II [12] we present the results of a further peer-evaluation study in which high-energy physicists were requested to distinguish between these two criteria in assessing the output of the world's six largest proton-accelerators.

[49] The DESY machines were, however, responsible for several lower-level but nonetheless major advances: the discovery of the decay of the psi prime to the psi via an intermediate state, and confirmation of the discoveries of the upsilon, upsilon prime and charmed F mesons. Moreover, since the end of 1978, PETRA has been responsible for several very important advances such as the discovery of 3-jet events.

[50] For example, CERN currently offers a greater range of facilities than Fermilab and Brookhaven combined.

the results based on highly cited papers are not completely satisfactory, since slightly different answers are obtained according to whether one focuses on "crucial discoveries", "major advances," or "advances." For a subject in the throes of revolutionary change, the first of these yields results that are most consistent with peer-evaluation. However, it is somewhat limited as an indicator because the statistics on crucial discoveries are small; while it may, therefore, be used to identify the leader(s) in the field, it is unlikely to discriminate clearly between the great majority of the research facilities being assessed.

From this, various conclusions can be drawn. First, it is not possible to carry out assessments of relative scientific outputs on the basis of a single bibliometric indicator (for example, publication or citation totals). To attempt to do so may lead to results that are seriously misleading. Second, even when several indicators are used together, one cannot adopt a "mechanical" approach whereby publication and citation data are computed and assumed to reveal in a faithful manner all aspects of the output from different research facilities. The results based on each indicator need to be carefully *interpreted*, as has been clearly demonstrated above.[51] Finally, it needs to be recognized that careful interpretation is only possible on the basis of detailed inside information on the research field under consideration, information that is accessible to "outsiders," such as policy-researchers, only through intensive interviewing of the type described earlier. Even if it is impossible to obtain quantitative peer-evaluation data (as we were unable to do for some aspects of experimental high-energy physics before 1969), qualitative data are still absolutely essential if one is to ensure that the various pitfalls associated with each of the bibliometric indicators are to be safely negotiated.

This concludes the first part of our assessment of the high-energy physics facilities operated by CERN. We have seen how the method of converging partial indicators can be used – admittedly only with great care – to yield what appear to be reasonable results comparing the overall output of the CERN accelerators with that of other major high-energy physics facilities over the period 1969–78. It seems clear that, over this period, users of the CERN accelerators contributed less to the advance of scientific knowledge than those of SLAC, but more than those of any other experimental centre. However, if we take into account the differing levels of resources invested in each accelerator laboratory (see tables 8 and 10), then the users of all three US national accelerator laboratories appear to have a better record over the ten years in question than those of CERN.[52] In paper II [12], we look more closely at the performance of the individual CERN accelerators, at the changes in their performance over time, especially in the period since 1978, and at the factors explaining their performance relative to that of other major accelerators around the world.

References

[1] E. Amaldi, CERN and the Big Machine, *Nature* 214 (1967) 1290–92.

[2] *CERN Annual Report 1977* (CERN, Geneva, 1977).

[3] *CERN Annual Report 1980* (CERN, Geneva, 1980).

[4] Meeting on Technology Arising from High-Energy Physics, *CERN Courier* 14 (1974) 111.

[5] 25th Anniversary of CERN, *CERN Courier* 19 (1979) 227.

[6] S. Coleman, The 1979 Nobel Prize in Physics, *Science* 206 (1979) 1290–92.

[7] M. Gibbons, The CERN 300 GeV Accelerator: A Case Study in the Application of the Weinberg Criteria, *Minerva* 8 (1970) 181–91.

[8] M. Goldsmith and E. Shaw, *Europe's Giant Accelerator* (Taylor and Francis, London, 1977).

[9] M.G.N. Hine, Industrial Implications of Large International Research Enterprises, *CERN Courier* 8 (1968) 174–83.

[10] J. Irvine and B.R. Martin, Assessing Basic Research: The Case of the Isaac Newton Telescope, *Social Studies of Science* 13 (1983) 48–86.

[11] J. Irvine and B.R. Martin, Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators, *Social Studies of Science* (forthcoming).

[12] J. Irvine and B.R. Martin, CERN: Past Performance and Future Prospects II. The Scientific Performance of the CERN Accelerators, *Research Policy* 13 (1984) (forthcoming).

---

[51] As a result, we would argue that there is little point in attempting to calculate correlation coefficients between the results based on different indicators – this would be more misleading than useful because it might suggest spurious precision.

[52] In paper II, we shall see how the situation has changed very significantly during the early 1980s, particularly as a result of the discoveries at CERN of the two Intermediate Vector Bosons.

AVAILABLE

[13] J. Irvine and B.R. Martin, What Direction for Basic Scientific Research?, in M. Gibbons, P. Gummett and B.M. Udgaonkar (eds.), *Science and Technology in the 1980s and Beyond* (Longman, London, 1984), pp. 67–98.

[14] K. Johnsen, The AGS and CERN PS: Recollections from the Early Years, in: N.V. Baggett (ed.), *AGS 20th Anniversary Celebration* (Brookhaven National Laboratory, BNL 51377, 1980) 34–40.

[15] W. Kirk, A Comparison of U.S. and Western European Funding for HEP (Stanford Linear Accelerator Center), mimeo (1978).

[16] T. Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press, Chicago, 1970).

[17] L.M. Lederman, The Upsilon Particle, *Scientific American* 239 (October 1978) 60–68.

[18] B.R. Martin and J. Irvine, Internal Criteria for Scientific Choice: An Evaluation of the Research Performance of Electron High-Energy Physics Accelerators, *Minerva* 19 (1981) 408–32.

[19] B.R. Martin and J. Irvine, Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy, *Research Policy* 12 (1983) 61–90.

[20] B.R. Martin and J. Irvine, CERN: Past Performance and Future Prospects. III. CERN and the Future of World High-Energy Physics, *Research Policy* (forthcoming)

[21] M.J. Moravcsik, The Crisis in Particle Physics, *Research Policy* 6 (1977) 78–107.

[22] European Council for Nuclear Research, *Nature* 170 (1952) 1104–05.

[23] An International Laboratory for Nuclear Research, *Nature* 172 (1953) 646–47.

[24] The Big Machine, *Nature* 214 (1967) 1283–84

[25] Britain, Europe and the Accelerator, *Nature* 219 (1968) 1196–97.

[26] What Next at CERN?, *Nature* 228 (1970) 593–94.

[27] Weathercock Turns Round, *Nature* 228 (1970) 1014–15.

[28] Particle Data Group, *An Indexed Compilation of Experimental High-Energy Physics Literature* (Lawrence Berkeley Laboratory, LBL-90, 1978).

[29] D.H. Perkins, Inelastic Lepton-Nuclear Scattering, *Reports on Progress in Physics* 40 (1977) 409–81.

[30] J. Polkinghorne, The Great Quark Hunt, *The Guardian* (10 July 1980) 13.

[31] C. Roche, Resources Given to High-Energy Physics in 1978 in CERN Member States (CERN, Geneva, DIR/CP0/239, 1979).

[32] H. Schmied, A Study of Economic Utility from CERN Contracts, *IEEE Transactions on Engineering Management* EM-24 (1977) 125–38.

[33] *Science Citation Index*, produced annually from 1961 to the present by the Institute for Scientific Information, Philadelphia.

[34] G. Trilling, *Report of the Subpanel on Long-Range Planning for the U.S. High-Energy Physics Program of the High-Energy Physics Advisory Panel* (Division of High-Energy Physics, U.S. Department of Energy, Washington, D.C., DOE/ER-0128, 1982).

[35] V.F. Weisskopf, ISR Inauguration 16 October 1971, *CERN Courier* 11 (1971) 291–92.

CERN: Past performance and future prospects

II. The scientific performance of the CERN accelerators

John IRVINE and Ben R. MARTIN *

Science Policy Research Unit, University of Sussex, Brighton BN1 9RF, UK

Final version received February 1984

In a series of three papers, we attempt to evaluate the past scientific performance of the three main particle accelerators at the Geneva-based European Organization for Nuclear Research (CERN) over the period since 1960, and to assess the future prospects for CERN and its users during the next ten to fifteen years.

We concerned ourselves in a previous paper (Paper I – Martin and Irvine [51] with the position of the CERN accelerators in world high-energy physics relative to those at other large laboratories working in the field. We dealt primarily with the period from 1969 to 1978, and attempted to establish how the experimental output from the three principal CERN accelerators, *taken as a whole*, compared with that from other major facilities. In undertaking this comparative evaluation, we drew on the method of "converging partial indicators" used in previous studies of three Big Science specialties.

In contrast, this paper (Paper II) focuses in detail on the scientific performance of each of the CERN accelerators *taken individually*. In particular, it asks, first, how the outputs from the CERN 28 GeV (giga or billion electron-volts) Proton Synchrotron compare with those from a very similar 33 GeV American accelerator at Brookhaven National Laboratory over the past two decades? Second, how great have been the experimental achievements of the Intersecting Storage Rings in world terms? And, third, how do the outputs from the CERN 400 GeV Super Proton Synchrotron and from a rival American machine at Fermi National Accelerator Laboratory compare? Attempts are then made to identify the main factors responsible for determining the relative scientific performance of each CERN machine.

These factors are of relevance to the subject of a third paper (Paper III – Martin and Irvine [52]) which sets out to assess the future prospects for CERN and in particular for LEP, the large electron–positron collider scheduled for completion in the latter part of the 1980s. What are the construction requirements (financial and technical) associated with LEP, and how easily will they be met? How does the scientific potential of LEP compare with that of other major accelerators under construction around the world? And, in the light of the previous record of the CERN accelerators, to what extent is this potential likely to be realized? The paper concludes with a discussion of the extent to which predictive techniques can be utilized in the formulation of scientific priorities, and of the problems in current science policy-making that such techniques might help address.

1. Introduction

Having summarized in Paper I (Martin and Irvine [51]) the methodology used to evaluate the past performance of the CERN accelerators, and obtained an overall picture of their position in world high-energy physics, we now turn to focus in more detail on the scientific outputs of the individual CERN machines.[1] The time-horizon taken in

---

* No order of seniority implied (rotating first authorship). The authors are Fellows of the Science Policy Research Unit, where they work on a range of issues connected with policies for basic and applied research. They gratefully acknowledge the support of the British Social Science Research Council in carrying out this research, and the help so freely given by large numbers of high-energy physicists. Thanks are also due to a number of colleagues at SPRU, and to the late Sir John Adams, Sir Clifford Butler, Professor John Dowell, Dr. Yves Goldschmidt-Clermont, Professor Tom Kibble, Dr. Leon Lederman, Dr. Owen Lock, and Dr. Paul Musset for providing useful critical comment on earlier drafts of this paper. However, the conclusions remain the responsibility of the authors alone.

[1] As in Paper I [51], it must be stressed that what is being assessed here is the scientific performance of the CERN accelerators and those who have used them, rather than CERN *per se*.

Research Policy 13 (1984) 247–284
North-Holland

0048-7333/84/$3.00 © 1984, Elsevier Science Publishers B.V. (North-Holland)

this paper is the period from 1961 [2] to 1983. However, given that this spans a number of qualitatively very different stages in CERN's history, it is convenient to break it down into a number of four-year periods. We shall concentrate in turn on each period, and, with the aid of various "partial indicators" based on bibliometric and peer-evaluation data, attempt to establish how successful each CERN accelerator has proved. Finally, using the results of interviews with a large number of high-energy physicists in North America, and Eastern and Western Europe, we shall discuss in detail the factors that appear to have determined the relative scientific contributions of the CERN accelerators and how these have changed over time.

2. 1961 to 1964: The commissioning and early work of the PS

1960 must have been a year of some excitement at CERN. The small 600 MeV (million electron-volts) synchro-cyclotron had been operating for three years, and had already claimed a notable discovery (the electron decay-mode of the charged pi-meson). It had also been used to carry out a pioneering "g-2" experiment (to measure the anomalous magnetic moment of the muon), the first in a series of ever more precise experiments spanning two decades for which CERN was to become famous within the world high-energy physics community. Moreover, the race to complete the construction of the 28 GeV Proton Synchrotron (PS) ahead of the rival 33 GeV Alternating Gradient Synchrotron (AGS) at Brookhaven National Laboratory on Long Island, New York, had been won by several months. [3] However, despite these early successes, the total research output from the PS during the period 1961 to 1964 was rather limited, as we shall see below. Perhaps the main problem was that, although the first experimental run on the PS took place in 1960, overall the preparations for experiments were generally still at a very early stage. Thus, while the PS was capable of accelerating particles and producing a beam, it was

> not yet ready to participate fully in research. This is why the first rich harvest in regard to these new questions was not reaped here. This is why the first quantum states, the first new particles, were in fact not discovered here (Weisskopf [73, p. 13]).

The full effects of not being ready to mount a comprehensive experimental programme can be seen in table 1. Certainly the PS began to yield a large number of research papers relatively quickly, accounting for 10 percent of the total world output of journal articles in the field during 1961–62, and nearly 30 percent in 1963–64. This was not only more than the Brookhaven AGS produced during this time (4.5 percent and 11 percent of the world output, respectively), but also more in the latter period than the well-established Bevatron accelerator at Berkeley in California (38 percent and 23 percent). However, the CERN papers reported data from relatively simple experiments, and consequently their overall impact appears to have been considerably less than that of publications from the other front-line proton accelerators. Table 1 shows that PS publications received 14.5 percent of all citations made in 1964 to experimental high-energy physics papers published between 1961 and 1964 – appreciably less than the AGS (23 percent) and the Bevatron (35 percent). The figures on average citations per paper show even more clearly the differences between the accelerators: PS publications earned an average of 1.9 citations per paper, which is very low for a new accelerator and compares rather unfavourably with the figure of 8.0 for the AGS (a typical value for a new accelerator) and that of 3.4 for the Bevatron (not unreasonable for an accelerator that had been operating for ten years). In addition, in terms of highly cited papers, the Bevatron (with 33 papers cited 15 or more times, and 5 cited 30 or more) and the AGS (with corresponding figures of 24 and 5) both seem to have been appreciably more successful than the CERN PS (with figures of 9 and 1, respectively). In his foreword to the *CERN*

---

[2] The very first papers from the CERN PS were in fact published in 1960. However, they were so few in number that their inclusion would make no difference to the results reported below. We chose to exclude them because no citation data are available in the *Science Citation Index* [62] for 1960.

[3] One of the main reasons for this was that Brookhaven built an electron analogue to test the new alternating-gradient technique. The CERN team apparently lacked the staff and funds needed for this. However, because of the close links that had been forged with Brookhaven, they were nevertheless able to learn almost as much from the American analogue as they would have from their own facility. Consequently, CERN was able to concentrate its efforts more immediately on the construction of the PS, giving it several months' head-start over Brookhaven, a lead which it maintained right until the accelerator was commissioned.

Table 1
Experimental high-energy physics, 1961–1964 [a]

	Numbers of experimental papers [b] published in past two years		Citations [c] to work of past four years	Average citations per paper	Highly cited papers (Number cited n times)			
	1962	1964	1964	1964	$n \geq 15$	$n \geq 30$	$n \geq 50$	$n \geq 100$
CERN PS (28 GeV)	35 (10.0%) [d]	155 (29.5%)	380 (14.5%)	1.9	9 (10.0%)	1 (6.0%)	0	0
Other W. Europe	30 (8.5%)	45 (8.0%)	130 (5.0%)	1.8	4 (4.5%)	1 (6.0%)	0	0
Total W. Europe	70 (18.0%)	200 (37.5%)	510 (19.5%)	1.9	13 (15.0%)	2 (12.0%)	0	0
Brookhaven AGS (33 GeV)	15 (4.5%)	60 (11.0%)	600 (23.0%)	8.0	24 (27.0%)	5 (29.5%)	1 (25.0%)	1 (100%)
Brookhaven Cosmotron (3 GeV)	45 (12.5%)	30 (5.0%)	280 (11.0%)	3.8	9 (10.0%)	4 (23.5%)	1 (25.0%)	0
Berkeley Bevatron (6 GeV)	145 (38.0%)	120 (23.0%)	910 (35.0%)	3.4	33 (37.5%)	5 (29.5%)	2 (50.0%)	0
Other US	25 (6.5%)	50 (9.0%)	210 (8.5%)	3.0	9 (10.0%)	1 (6.0%)	0	0
Total US	230 (61.5%)	255 (48.0%)	2000 (77.5%)	4.1	75 (85.0%)	15 (88.0%)	4 (100%)	1 (100%)
Dubna (10 GeV)	70 (18.5%)	65 (12.0%)	70 (3.0%)	0.5	0	0	0	0
Moscow ITEP (7 GeV)	5 (2.0%)	15 (3.0%)	10 (0.5%)	0.5	0	0	0	0
Other – rest of world	–	–	–	–	–	–	–	–
Total – rest of world	75 (20.5%)	80 (15.0%)	80 (3.0%)	0.5	0	0	0	0
World total	375 (100%)	535 (100%)	2590 (100%)	2.8	88 (100%)	17 (100%)	4 (100%)	1 (100%)

[a] The notes to tables 7 and 9 in Martin and Irvine [51] also apply.
[b] All publication figures have been rounded to the nearest 5.
[c] All citation figures have been rounded to the nearest 10.
[d] All percentage figures have been rounded to the nearest 0.5%.

Annual Report 1965, J.H. Bannier, then President of the CERN Council, referred to certain decisions which

> will enable CERN to maintain its position as one of the world's three leading laboratories in the field (Bannier [4, p. 3]),

and these various bibliometric indicators would suggest that, at the end of 1964, the CERN PS did indeed rank in third position, signficantly behind the Berkeley Bevatron and the Brookhaven AGS accelerators.

In opening up a new energy range in high-energy physics, experience shows that there are often a number of discoveries to be made, some predicted in advance and others completely unexpected. The data on highly cited papers suggest that most of the discoveries in this particular energy range were made not by the PS but by the Brookhaven AGS, as the CERN Director-General himself admitted at the time:

> Certainly, the results could have been more impressive. Some of the recent sensational discoveries were made elsewhere (Weisskopf [73, p. 18]).

Three discoveries at the AGS stand out in particular:

(1) the discovery that there are two different types of neutrinos;
(2) the discovery of the omega-minus particle;
(3) the discovery of charge-parity or CP-violation.

It is worth examining the histories of these three discoveries since they reveal much that can be used to explain CERN's record during the early 1960s.[4]

The idea of undertaking an experiment to determine whether the electron-neutrino and the muon-neutrino are the same or different particles seems to have occurred more or less simultaneously to experimenters at Brookhaven and CERN in about 1960. It was clear that the higher energies of the AGS and PS would, for the first time, permit neutrino experiments to be carried out on a particle accelerator, and the two-neutrino experiment was, as one of the physicists involved has pointed out (Schwartz [61, p. 42]), *the* obvious experiment to do with neutrinos.[5] Such an experiment was attempted at CERN in 1961, but without success (Weisskopf [71, p. 16]). As the Director-General of the time later observed:

> we tried early to construct a beam of neutrinos – perhaps too early – and the first experiment did not lead to any results because of lack of intensity (Weisskopf [73, p. 16]).

and in subsequent years, this failure was attributed by the director of one of the CERN Departments to

> a still inhomogeneous and inexperienced group … [who] made some mistakes in evaluating the fluxes of particles available (Cocconi [34, p. 349]).

One of those involved in the rival American experiment recollected his experience of the situation as follows:

> [The PS] had a head start on us at CERN, and it was clear that they would be at least six months ahead of us, maybe even a year … [T]he question then was: Would CERN beat us? And indeed the entire experiment was designed at CERN and was almost ready to mount and then we heard … that Von Dardel had discovered a mistake in the calculations and indeed it turned out that the beam as it was planned for CERN would give very low intensity because the beam was planned for a 5-foot straight section and the defocusing effect of the magnets right after the straight section would have essentially demolished the beam … The other mistake, which is in a sense much deeper and is really an indication of the very real difference in physics philosophy in [the U.S.] compared to what it was there [at CERN], was that rather than saying this experiment is so important let's move it to the 10-foot straight section, the hell with all the other junky little experiments that are going on, they cancelled the experiment, and of course we knew they would because that was just the style (Schwartz [61, p. 45]).

The clear implication is that, if a similar mistake had been made at Brookhaven, it would have rapidly been rectified by moving less important experiments out of the way. This would not have been too difficult to arrange since

> there were no committees to decide whether you run or not. It was just Maurice [Goldhaber, the Laboratory Director] and … his great wisdom prevailed [*ibid.*].

The history of the second major discovery, that of the "omega-minus" particle, also suggests a significant difference in experimental approach between Brookhaven and CERN. The existence of the omega-minus had been predicted in 1962 by theorists who showed that in the SU(3) classification scheme – which successfully accounted for all the known hadron resonances – there was a "gap" corresponding to a new, previously unobserved, particle. The subsequent search for the predicted omega-minus began at CERN and Brookhaven towards the end of 1962, and again assumed the form of a "race" between the user-groups of the two rival accelerators. At CERN, experimenters initially used a beam of K mesons of momentum 3.5 GeV/c, but found no sign of the particle. They therefore decided to raise the K^- momentum to 6 GeV/c,

> but, due to a number of difficulties, this experiment was delayed as late as the summer of 1964. Meanwhile the Brookhaven group set up a 5 GeV/c K^- beam and early in 1964 they found a truly remarkable photograph of a Ω^- production and decay … This was a great triumph for the new [SU(3)] theory involving particle symmetries (Butler [8, p. 767]).

One of the major difficulties appears to have

---

[4] In discussing experimental work here and elsewhere in the paper, we have inevitably been forced to resort to a certain amount of high-energy physics "jargon" in giving names of particles, energy parameters, details of equipment, and so on. We have attempted to keep this to a minimum. The interested reader is referred to articles such as those by Georgi [37], Jacob and Landshoff [45], Polkinghorne [56], Salam [59], 't Hooft [42], and Weinberg [70] for recent popular commentaries on high-energy physics.

[5] Theory suggested that the only way to explain the apparent failure of the muon to decay into an electron plus a gamma ray was to posit the existence of two different types of neutrinos (Schwartz [61, p. 42]).

BEST COPY AVAILABLE

involved the British 1.5-metre bubble-chamber that was to be used as the particle detector in this experiment. It was reported in 1959 that this would be ready for installation at CERN during 1961 [54, p. 944]. In the event, its arrival from Britain was delayed until 1963, and it then failed to pass its initial operational trials on the PS in the early part of 1964 [9, p. 76]. The first experimental run with the bubble-chamber and the high-energy K^- beam was therefore delayed until June 1964.[6]

The third and probably most important of the early Brookhaven AGS discoveries was that of CP-violation.[7] Two points should be noted in connection with the planning and execution of this experiment. The first is that the experimental proposal was only one and a half pages long, and it freely admited that no detailed calculations had yet been made (cf. Fitch [35, pp. 49–51]). Whereas such a proposal was acceptable at Brookhaven, where decisions on experiments were generally made by the Director alone, its chances of negotiating the formal committee system at CERN would probably have been somewhat slimmer. Instead of trusting the intuition of the two principal physicists involved (both were highly respected experimenters), a committee would almost certainly have insisted that the calculations were first made, thus delaying the experiment. In the event, the experimenters at Brookhaven were able to start taking data within a mere two months of the proposal being written. A second point to note is that a suitable detector already existed and this further shortened the time needed to mount the experiment. This detector had previously been used for an experiment on the lower-energy Cosmotron accelerator at Brookhaven, and therefore could be moved very quickly to the AGS. Indeed, the detector was installed *before* the experiment was approved (cf. Robinson [58, p. 620]), again illustrating the relatively informal system for planning and executing experiments then prevalent at Brookhaven. The fact that the 3 GeV Cosmotron had been operational at Brookhaven for the previous ten years obviously gave the laboratory a considerable advantage over CERN.

It is clear that one of the main problems at CERN during this early period was the lack of experimental equipment. Nowhere was this more apparent than in relation to bubble-chambers which, it was beginning to be recognized, had clear advantages in many areas of high-energy physics over more traditional experimental techniques (such as emulsions or cloud-chambers). As a result, a great deal of effort was directed in the United States to the development of this new technology. Initially, Europe failed to match this pace of development, as the Director-General of CERN admitted in 1963:

> CERN has lagged behind the American laboratories with respect to the use of big bubble-chambers (Weisskopf [72, p. 13]).

At the time of the first experiments on the CERN PS in 1960, the largest bubble-chamber at the laboratory was a mere 300 cm (12 inches) in size, compared with the 72-inch instrument in use at the Berkeley Bevatron accelerator and the 80-inch chamber very shortly to become operational at the AGS. The situation improved at CERN in 1961 when the CERN 1-metre propane bubble-chamber was completed and a French user-group brought an 81-cm hydrogen chamber from Saclay. However, as we have seen, the large British 1.5-metre bubble-chamber did not start running experiments until the summer of 1964, three years after originally planned. Similarly, the 2-metre hydrogen bubble-chamber under construction at CERN was not completed until December 1964, having been delayed two years by staff shortages (cf. Weisskopf [71, p. 18]).

Perhaps the other main problem at CERN during these crucial early years of the PS was "a somewhat ill-defined experimental physics programme" [13, p. 345]). In 1962, the Director-General referred to "the haphazard and timid approach which characterized the beginning of research work" (Weisskopf [71, p. 16]), while another senior CERN official later described the preparation for the experimental programme as "abnormally poor" (Cocconi [34, p. 349]). Four main factors can be seen as underlying these problems with the physics programme and the late provision of appropriate experimental equipment at the laboratory.

---

[6] Another possible reason for the delay in using the higher-energy K^- beam was that priority was given to a search for the omega-minus using the antiproton beam from the PS. This search proved unsuccessful because the antiproton cross-sections turned out to be lower than predicted.

[7] This was the only high-energy physics experiment of the 1960s to be honoured with the award of a Nobel Prize.

First, there is the question of the overall management of the PS construction programme. In particular, it could be argued that senior CERN officials placed too much emphasis on early completion of the accelerator (preferably ahead of Brookhaven), and too little on developing a suitable experimental programme (and the construction of the necessary chambers and detectors). Obviously, opinions may differ as to whether this was a primary cause of the accelerator's early under-exploitation, but the subsequent observation by Cocconi that

> most of the activity [at CERN] had been devoted to the construction of the accelerator itself (*ibid.*, p. 349)

suggests that he for one had doubts about whether the balance between construction and experimental preparations had been optimal.

Second, it seems that CERN's failure to appreciate fully the need for appropriate experimental instrumentation was related to the comparative lack of experience of European experimentalists at that time. Their American counterparts had been carrying out research on the large accelerators at Brookhaven and Berkeley since the early 1950s, and therefore knew full well what was involved in mounting "state-of-the art" experiments. In particular, their greater experience meant that they were better able to adapt their instruments to new physics needs as soon as they arose, rather than having to wait for sophisticated purpose-built instruments to be constructed. It is probably not without significance that what was arguably the most important experiment to be carried out on the PS in the early 1960s (a study of the decay patterns of sigma hyperons – this yielded the most highly cited publication from the accelerator between 1961 and 1964) was carried out by a collaboration which included an American group.

Third, as in any new multinational research venture, it inevitably took time for an effective organizational structure to develop within CERN, particular problems arising in defining the role of the laboratory *vis-à-vis* user-groups. The original idea was that, while CERN was responsible for the machine, the beams and detectors would be the responsibility of individual groups, with "national truck teams," composed of university physicists from each Member State, arriving at the laboratory with the equipment needed to carry out their experiments. Under this approach, CERN would be expected to provide only a very limited range of experimental equipment for outside user-groups. This may explain why in the early years most of the emphasis at CERN seemed to be on the construction of the accelerator, with far less on beam equipment,[8] and very little on the provision of subsidiary experimental instrumentation and the planning of a coherent and systematic experimental programme that integrated the research activities of central staff and outside users and permitted full exploitation of the new facilities. Very quickly, however, it was found that the "truck team" approach "did not work out successfully" (Adams [2, p. 86]). This necessitated the laboratory having to develop a better integrated research policy, with experimental teams being brought together more by their common interest in a particular experiment and by the existence of complementary technical skills than by their nationality. Not only did this result in a greater internationalization of effort, but it also imposed heavier demands on the laboratory in providing both administrative and technical support. Inevitably, CERN took time to adjust to these greater demands placed upon it.

A fourth and final factor contributing to the problems of the PS during its early years of operation was the commitment of significant sections of the European high-energy physics community to existing research programmes on national accelerators. Three of the four large Member States – Britain, France and Italy – were each operating (or in the process of building) two national accelerator laboratories, while the fourth – West Germany – had one major centre. Consequently, many European physicists were engaged in planning or running experiments on these accelerators, with much of the available technical support channelled into constructing the necessary detectors and instrumentation. (Moreover, in the United Kingdom, there had been a total lack of co-ordination between the domestic experimental efforts[9] and those at CERN, at least until the early 1960s

---

[8] There were, for example, severe difficulties with separated beams in the early 1960s. In addition, there was initially a tendency to distinguish work into the categories of "accelerator," "beam," and "detector," with the result that these three elements did not always work well together.

[9] For an evaluation of the scientific outputs from one of the British national laboratories (Daresbury), see Martin and Irvine [50].

(cf. Wilkinson [74, pp. 8–19]) – hardly the best way of trying to ensure that the available resources were utilized as effectively as possible.) Perhaps with greater support from within its Member States, CERN might have been able to mount a comprehensive experimental programme on the PS somewhat sooner than it did. Further discussion of the factors affecting the scientific performance of the PS will, however, be left until later when we have seen how it fared in subsequent years.

3. 1965 to 1968: The CERN PS comes of age

CERN and the PS accelerator entered 1965 in a better prepared state than had been the case four years earlier. In particular, the accelerator had since 1963 been providing a neutrino beam with 100 times the intensity of that originally available on the Brookhaven AGS. Equally important, the new 2-metre bubble-chamber had finally come into operation at the end of 1964. Last, CERN had begun in 1964 to plan a PS Improvements Programme, the aim being to increase the intensity and repetition-rate of the accelerator and to extend the experimental areas. The first effects of Phase I of this programme began to be felt in the latter half of the four-year period under consideration. To what extent was this greater state of preparedness reflected in the scientific outputs from the PS over the four years up to 1968?

Table 2 shows that during this period the PS accounted for just under 30 percent of all experimental high-energy physics papers, considerably more than the Brookhaven AGS (19.5 percent) and indeed all other accelerators in the world. However, the total numbers of citations for the papers from the AGS and PS were virtually identical (about 27.5 percent of the world total). This suggests that the two machines had a very similar overall impact on the advance of high-energy physics during the period, and significantly more than their nearest competitor, the Berkeley Bevatron. The Bevatron's world share of publications and citations was by now declining rapidly because of its low energy relative to the PS and AGS, although it was still well ahead of newer machines of similar or slightly higher energy at Rutherford and Argonne.

In terms of "advances" – that is papers cited 15 or more times in one year – the PS had managed to catch up and perhaps even overtake the AGS, producing 46 such papers compared with 42. It had been particularly successful in discovering several new resonances, and in investigating the properties of others more thoroughly than had previously been achieved. Yet the more important "major advances" (papers cited 30 or more times in a year) still eluded the PS – it had only one such paper to its credit during the four years compared with 12 for the AGS. Moreover, some of the results reported in this single highly cited CERN paper were subsequently shown to be "mistaken." These results were obtained using a missing-mass spectrometer, and appeared to show a double-peak structure for the A_2 resonance. This was the very first indication of the "split A_2," as it came to be known. The finding was "confirmed" in several other papers by the same collaboration, by no less than four other collaborations working on the PS, and even by a team at the Brookhaven AGS. By the time of the major high-energy physics conference of 1969 (at Lund), it was accepted that

> everyone is now agreed that the A_2 is split [12, p. 233]

It was only in 1970 that experiments at the Stanford Linear Accelerator Center (SLAC), and later at Brookhaven and CERN, began to indicate that the A_2 was not split after all. By the time of the major conference of 1972 (held at Chicago), it was reported that the split A_2

> now seems to have been definitely swept under the carpet [15, p. 316]

All this is not to imply that the split A_2 was the only "mistake" of experimental high-energy physics during the 1960s. Far from it. Most of the major accelerators had at least one "mistake" to their credit. In the case of the CERN PS, there were Maglic's S, T and U resonances that other experimenters found great difficulty in corroborating; there was also the early high value for $|\eta_{00}|$, a result "confirmed" by an experiment at about the same time on the Princeton-Pennsylvania Accelerator. At Berkeley, the Bevatron gave an erroneously high value for the $(K_L - K_s)$ mass difference, as well as results that apparently violated the $\Delta S = \Delta Q$ rule. Finally, at Brookhaven, one AGS experiment claimed to find C-violation in the decay of the η particle. However, the split A_2 was by far the longest-lived mistake. Whereas the others

Table 2
Experimental high-energy physics, 1965–1968 [a]

	Numbers of experimental papers [b] published in past two years		Citations [c] to work of past four years		Average citations per paper		Highly cited papers (Number cited n times)			
	1966	1968	1966	1968	1966	1968	n ≥ 15	n ≥ 30	n ≥ 50	n ≥ 100
CERN PS (28 GeV)	210 (32.5%) [d]	215 (25.5%)	1280 (28.5%)	1330 (26.0%)	3.5	3.1	46 (30.5%)	1 (5.5%)	0	0
DESY (7 GeV)	10 (1.5%)	40 (4.5%)	20 (0.5%)	300 (6.0%)	1.6	6.1	10 (6.5%)	1 (5.5%)	0	0
Rutherford Nimrod (7 GeV)	15 (2.5%)	20 (2.5%)	100 (2.5%)	180 (3.5%)	6.0	4.6	9 (6.0%)	0	0	0
Other W. Europe	50 (7.5%)	50 (5.5%)	220 (5.0%)	190 (3.5%)	2.4	1.9	2 (1.5%)	0	0	0
Total W. Europe	290 (44.5%)	325 (38.5%)	1620 (36.0%)	1990 (39.0%)	3.3	3.3	67 (44.5%)	2 (11.0%)	0	0
Argonne ZGS (12 GeV)	15 (2.0%)	55 (6.5%)	50 (1.0%)	250 (5.0%)	4.2	3.6	6 (4.0%)	0	0	0
Brookhaven AGS (33 GeV)	125 (19.5%)	165 (19.5%)	1270 (28.0%)	1350 (26.5%)	6.8	4.6	42 (28.0%)	12 (66.5%)	1	0 (100%)
Brookhaven Cosmotron (3 GeV)	30 (4.5%)	20 (2.5%)	200 (4.5%)	160 (3.0%)	1.6	3.2	3 (2.0%)	0	0	0
Berkeley Bevatron (6 GeV)	80 (12.5%)	95 (11.5%)	780 (17.5%)	500 (10.0%)	3.8	2.8	9 (6.0%)	0	0	0
CEA (6 GeV)	15 (2.5%)	35 (4.0%)	150 (3.0%)	200 (4.0%)	5.4	3.9	7 (4.5%)	0	0	0
SLAC (20 GeV)	–	20 (2.5%)	–	100 (2.0%)	–	4.5	4 (2.5%)	1 (5.5%)	0	0
Other US	35 (5.0%)	60 (7.0%)	260 (6.0%)	410 (8.0%)	3.7	4.4	11 (7.5%)	2 (11.0%)	0	0
Total US	300 (46.5%)	450 (53.5%)	2710 (60.0%)	2950 (58.0%)	4.9	3.9	82 (54.5%)	15 (83.5%)	1	0 (100%)
Dubna (10 GeV)	40 (6.5%)	35 (4.0%)	120 (2.5%)	60 (1.0%)	1.1	0.9	0	0	0	0
Moscow ITEP (7 GeV)	15 (2.5%)	25 (3.0%)	50 (1.0%)	30 (0.5%)	1.7	0.6	0	0	0	0
Other – rest of world	–	10 (1.0%)	–	50 (1.0%)	–	4.2	1 (0.5%)	1	0	0
Total – rest of world	60 (9.0%)	70 (8.5%)	170 (3.5%)	140 (2.5%)	1.2	1.1	1 (0.5%)	1 (5.5%)	0	0
World total	645 (100%)	845 (100%)	4500 (100%)	5080 (100%)	3.8	3.4	150 (100%)	18 (100%)	1	0 (100%)

[a] The notes to tables 7 and 9 of Martin and Irvine [51] also apply.
[b] All publication figures have been rounded to the nearest 5.
[c] All citation figures have been rounded to the nearest 10.
[d] All percentage figures have been rounded to the nearest 0.5%.

were corrected within two to three years, the split A_2 took five or six years to rectify. During this time, it caused "consternation" to theorists (Jentschke [46, p. 23]), occupying a considerable part of their attention because it cast doubt on the otherwise very successful SU(3) classification scheme for hadrons. The high number of citations earned by the original split A_2 paper suggests that at the time it had a major impact on the scientific community, stimulating much experimental and theoretical work that might not otherwise have been carried out. However, as with the "mistaken"

Fermilab papers discussed in Paper I (Martin and Irvine [51, p. 201]), it is debatable whether such a publication can be regarded as a positive contribution to the advance of scientific knowledge.

To sum up, although the PS had a higher publication output than the AGS over the four-year period up to 1968, papers from the two machines earned a similar total number of citations. However, the main difference between the two accelerators was that the AGS had several "major advances" to its credit while the PS had only one (and one which was later to prove of somewhat dubious merit). In terms of these bibliometric indicators, then, the overall scientific contribution of the PS would appear to have been smaller than that of the AGS [10], but somewhat greater than both the Berkeley Bevatron and the new accelerators at DESY and Argonne.

What reasons can be given for the apparently superior record of the Brookhaven AGS over this period? First, although the difference was less pronounced than hitherto, users of the CERN PS continued on average to be less experienced in using large accelerators than their Brookhaven counterparts. Second, the AGS was by all accounts the better instrumented of the two accelerators; the CERN PS was still several years from obtaining the very large 25-cubic metres hydrogen bubble-chamber, and the 10-cubic metres heavy-liquid bubble-chamber, recommended by the Amaldi Committee (the forerunner of the European Committee for Future Accelerators) as far back as 1963. Finally, the attention of at least some CERN staff was beginning to switch to the Intersecting Storage Rings project and to the even more distant 300 GeV (as it then was) Super Proton Synchrotron. Brookhaven staff, by contrast, had no such major new facility to which they could look forward.[11] However, as with our previous consideration of the period 1961–64, we shall leave further discussion of the factors determining the performance of the PS until later.

4. 1969 to 1972: CERN achieves world pre-eminence

As mentioned previously, the Amaldi Committee had in 1963 recommended a major 200 million Swiss franc programme at CERN to: (1) build large hydrogen and heavy-liquid bubble-chambers together with a range of other detectors; (2) improve the technical performance of the Proton Synchrotron; and (3) increase the number of experimental halls. Although Phase I of the PS Improvement Programme had been completed in 1967, it was not until 1969 that the effects of Phase II began to be reflected in a considerably increased accelerator intensity (with a mean intensity of 1.3×10^{12} protons per pulse for the year, nearly twice the figure for 1965, and ten times greater than that for 1961 – see [10, p. 77]). Then, in 1971, the new heavy-liquid bubble-chamber finally came into operation. These two major developments together constituted a considerable upgrading of the PS facility, and for perhaps the first time gave CERN user-groups a significant technical advantage over physicists undertaking experimental work on the Brookhaven AGS. Moreover, 1971 also saw the completion of the Intersecting Storage Rings (ISR), within the allocated budget and six months ahead of schedule, as well as the go-ahead (after several years of complex intergovernmental negotiations) for construction of the Super Proton Synchrotron. All this gave rise to "a state of unaccustomed euphoria" at the laboratory [14, p. 31]. The first observation of interactions at the ISR was described in the following terms:

---

[10] Even at CERN, this conclusion would probably not now be disputed, as this commentator's assessment makes clear: "For a number of years in the 1960s, Brookhaven was the finest high-energy physics laboratory in the world." [26, p. 247]. However, whether this assessment would have been accepted in 1968 is less obvious. At the end of that year, L. Van Hove (formerly Head of the Theoretical Physics Department at CERN and a future Director-General) undertook a review of recent contributions from CERN accelerators to high-energy physics. He identified five major contributions, one on the synchro-cyclotron (the latest g-2 experiment) and four on the PS: (1) leptonic decays of vector mesons; (2) the discovery of several new mesons; (3) the split A_2; and (4) tests of Cabibbo theory. It was somewhat unfortunate that, of the four contributions from the PS, the split A_2 should later be found to be "mistaken," while several of the new resonances (the S, T and U mesons) subsequently came under considerable doubt.

[11] Brookhaven did put itself forward as a possible site for the planned American 200 GeV proton synchroton which eventually went to the Mid-West (Fermilab). However, the involvement of Brookhaven staff in this project was slight, the design-study having been carried out at Berkeley (another unsuccessful contender for the site).

> Never before in accelerator history has commissioning of a machine gone so smoothly, despite the fact that the ISR is the most complex machine ever built and, in the sense of doing something new whose successful operation could not be guaranteed, is among the most adventurous ever built. To be with the jubilant physicists that day when they were clocking up particles coming from interactions in the colliding beams was a rare experience. Not every day does excitement break through the surface so openly [14].

To what extent was this mood of optimism about the future for the ISR subsequently matched by significant experimental outputs from the facility? And what effect did the Improvement Programme have on the scientific contributions made by the PS at the start of the 1970s?

Data on the scientific publications arising from the world's main accelerators over the period 1969–72 are given in table 3. As in 1968, the CERN PS continued to produce about 25 percent of the world total of experimental publications. This was a considerably greater proportion than that achieved by the Brookhaven AGS, whose share dropped from 19 percent to 14.5 percent in 1972, principally because a major conversion programme [12] for the accelerator proved rather more troublesome than that undertaken at the PS and required the total shut-down of the facility for an extended period. The effects of this can be seen very clearly in the citation data. While papers for the two accelerators earned virtually the same number of citations in 1970, the AGS figure had by 1972 dropped appreciably below that of the PS, largely because of the decline in the AGS publication rate (AGS papers were still gaining more citations per paper – 2.9 compared with 2.3 for the PS). The figures on highly cited papers were also very similar, although both accelerators managed only one "major advance" (papers cited 30 or more times in a year) between 1969 and 1972. Overall, it would seem that the PS had finally drawn level with the AGS (it had been appreciably behind in the previous four-year period in terms of "major advances"), and perhaps had even begun to edge slightly ahead.

However, for this period, the most highly cited papers came neither from the PS nor the AGS, but from other newer accelerators. In particular at Serpukhov, the first indications of a rising total cross-section in hadron collisions were found by a joint Soviet–CERN experiment. The discovery of deep inelastic scattering was made by researchers at the Stanford Linear Accelerator Center; this result, presented in the two SLAC papers [13] cited 50 or more times in a year, was important because it provided the first direct evidence that protons were composed of smaller constituent "partons" (subsequently identified with "quarks"). The very first results from Fermilab on proton–proton total cross-sections were published in the latter part of 1972, and these too were highly cited. And, finally, there were two [14] important early results from the ISR: (1) measurements of inclusive reactions, which showed the same "scaling" behaviour as had first been seen in the SLAC deep inelastic scattering experiments; and (2) the discovery of the diffraction minimum in elastic proton–proton scattering, contradicting the predictions of the previously fashionable Regge theory (Jentschke [46, pp. 20–22]). It is evident from table 3 that the SLAC, Serpukhov and ISR machines all achieved a higher impact in terms of citations per paper than either the PS or AGS. Indeed, given the relatively high publication output of SLAC over this period, it is probably fair to conclude that it ranked as the most successful accelerator over this period, closely followed by Serpukhov and the ISR. With SLAC taking over top position from the Brookhaven AGS, the Americans thus retained their record of operating the accelerator contributing most to the advance of high-energy physics. However, if the contributions from the ISR and PS are taken together, then the bibliometric data suggest that the CERN research facilities contributed more to the field than those at any other laboratory in the world. CERN was probably followed by SLAC, Serpukhov, and Brookhaven, with the Berkeley and Argonne accelerators sharing fifth place, a little ahead of DESY. Thus, just over a decade after the PS had come into operation, CERN finally achieved the distinction of becom-

---

[12] As with the CERN PS Improvement Programme, the aim was to increase the intensity of the proton beam. However, the problems experienced at Brookhaven with the conversion programme were so great that "the AGS operated initially with beam intensities somewhat lower than pre-shutdown values" [7, p. 47].

[13] These papers were published in 1969, although the experiment had actually been carried out in 1968.

[14] The discoveries of high transverse-momentum events and of rising total cross-sections for proton–proton collisions were made in 1972, but, since the results were not published until 1973, they are considered in the discussion of the period 1973–76 in the next section.

Table 3
Experimental high-energy physics, 1969–1972 [a]

	Numbers of experimental papers [b] published in past two years		Citations [c] to work of past four years		Average citations per paper		Highly cited papers (Number cited n times)			
	1970	1972	1970	1972	1970	1972	$n \geq 15$	$n \geq 30$	$n \geq 50$	$n \geq 100$
CERN ISR (28 + 28 GeV)	–	20 (2.0%) [d]	–	270 (5.0%)	–	12.9	13 (10.5%)	4 (19.0%)	2 (22.0%)	0
CERN PS (28 GeV)	240 (25.0%)	255 (25.0%)	1200 (22.5%)	1150 (21.5%)	2.6	2.3	19 (15.5%)	1 (5.0%)	0	0
DESY (7 GeV)	25 (2.5%)	35 (3.5%)	280 (5.5%)	250 (4.5%)	4.4	4.0	3 (2.5%)	0	0	0
Rutherford Nimrod (7 GeV)	25 (2.5%)	25 (2.5%)	140 (2.5%)	100 (2.0%)	3.2	1.9	0	0	0	0
Daresbury NINA (4 GeV)	10 (1.0%)	20 (2.0%)	20 (0.5%)	70 (1.5%)	2.1	2.7	0	0	0	0
Other W. Europe	45 (5.0%)	45 (4.5%)	240 (4.5%)	250 (4.5%)	2.5	2.7	6 (5.0%)	2 (9.5%)	0	0
Total W. Europe	340 (36.0%)	405 (39.5%)	1870 (35.5%)	2090 (39.5%)	2.8	2.8	41 (33.5%)	7 (33.5%)	2 (22.0%)	0
Argonne ZGS (12 GeV)	90 (9.5%)	95 (9.5%)	400 (7.5%)	490 (9.0%)	2.8	2.6	7 (5.5%)	0	0	0
Brookhaven AGS (33 GeV)	180 (19.0%)	150 (14.5%)	1200 (23.0%)	960 (18.0%)	3.5	2.9	22 (18.0%)	1 (5.0%)	0	0
Berkeley Bevatron (6 GeV)	85 (9.0%)	60 (6.0%)	520 (10.0%)	350 (6.5%)	2.9	2.5	8 (6.5%)	1 (5.0%)	0	0
CEA (6 GeV)	20 (2.5%)	10 (1.0%)	170 (3.0%)	110 (2.0%)	3.0	3.3	2 (1.5%)	0	0	0
Cornell (12 GeV)	15 (1.5%)	15 (1.5%)	130 (2.5%)	70 (1.5%)	4.8	2.2	4 (3.5%)	0	0	0
SLAC (20 GeV)	55 (5.5%)	80 (8.0%)	460 (8.5%)	670 (12.5%)	6.1	4.9	21 (17.0%)	4 (19.0%)	2 (22.0%)	0
Other US	50 (5.0%)	30 (3.0%)	270 (5.0%)	150 (3.0%)	2.5	1.9	6 (5.0%)	4 (19.0%)	3 (33.5%)	0
Total US	490 (51.5%)	445 (43.5%)	3140 (59.5%)	2800 (53.0%)	3.3	3.0	70 (57.0%)	10 (47.5%)	5 (55.5%)	0
Serpukhov (70 GeV)	30 (3.0%)	60 (6.0%)	140 (2.5%)	270 (5.0%)	4.8 (10.7) [e]	3.0 (6.4) [e]	12 (9.5%)	4 (19.0%)	2 (22.0%)	1 (100%)
Other – rest of world	90 (9.5%)	110 (11.0%)	120 (2.0%)	140 (2.5%)	0.7	0.5	0	0	0	0
Total – rest of world	120 (12.5%)	170 (17.0%)	260 (5.0%)	420 (8.0%)	1.4	1.4	12 (9.5%)	4 (19.0%)	2 (22.0%)	1 (100%)
World total	950 (100%)	1020 (100%)	5270 (100%)	5310 (100%)	2.9	2.7	123 (100%)	21 (100%)	9 (100%)	1 (100%)

[a] The notes to tables 7 and 9 of Martin and Irvine [51] also apply.
[b] All publication figures have been rounded to the nearest 5.
[c] All citation figures have been rounded to the nearest 10.
[d] All percentage figures have been rounded to the nearest 0.5%.
[e] This is the figure for papers in Western journals only.

ing the world's foremost high-energy physics laboratory. This elevation into a position of pre-eminence was clearly sensed at CERN, as the following quotations from the *1972 Annual Report* make apparent:

> The results from European physicists, working at CERN or

> using its facilities, were recognized to be of outstanding quality at the International Conference on High-Energy Physics which took place at Chicago. Many scientists at the Conference acknowledged that, in its field, the CERN Laboratory is second to none (Jentschke [47, p. 11]).

> For particle physics at CERN, 1972 proved to be an outstanding year, the pre-eminent place of the Organization's research being generally recognized, in particular at the Batavia conference. Not only was the PS physics programme at least as interesting as in the past but also there were the first exciting results from the ISR [11, p. 26].

5. 1973 to 1976: The American revolution

To what extent were CERN users able to maintain and build upon the pre-eminent position achieved in 1972? On the technical side, various developments took place that served to strengthen the laboratory's research capability. On the Proton Synchrotron, a new Booster Injector permitted greatly increased intensities of particle beam to be obtained. This was to prove of crucial importance for the investigation of weak interactions, in particular those involving neutrinos, enabling the full potential of the Gargamelle heavy-liquid bubble-chamber to be exploited.[15] In addition, the major new Omega Spectrometer came into operation in 1973, being used in PS experiments from 1973 to 1975 and (with some modification) on the Super Proton Synchrotron (SPS) thereafter. However, one disappointment was that the very large hydrogen bubble-chamber (BEBC – the Big European Bubble-Chamber) could only be used for a few experiments on the PS in 1973[16] and again in 1975. Virtually the whole of 1974 was lost when the BEBC magnet had to be completely dismantled to rectify a short-circuiting fault. As a result, this major experimental facility – first recommended in 1963 as part of the programme to exploit the PS – was in the event only used for a short time on the accelerator (although, as we shall see later, it proved invaluable in the SPS experimental programme). Besides these improvements in the research capability of the PS, the Intersecting Storage Rings (ISR) were by now coming into their own as a powerful experimental facility, partly because of continual increases achieved in luminosity (and hence in the number of collisions per second), and partly because of the large Split Field Magnet detector which began operating in 1973.

With the benefit of all these technical advances, CERN witnessed in 1973 what was probably the most successful year in its history (at least until 1983). Not only did European researchers make the fundamental discovery of neutral currents with the Gargamelle bubble-chamber on the PS, but several other major advances also arose from experiments on both the PS and ISR. (The details are discussed below.) Thus, at the end of the year, the Director-General was able to report with some pride that,

> In 1973 CERN led major advances in our knowledge of the fundamental properties of matter and of the forces determining the behaviour of matter in our Universe (Jentschke [48, p. 11]).

However, the period from 1973 to 1976 was one of the most tumultuous in the history of high-energy physics, ushering in the revolutionary era of what became known as the "new physics." Despite the promising start made by CERN users in 1973, the central role in these developments was subsequently filled by accelerators elsewhere, as can be seen from table 4. True, the CERN PS and ISR were still able between them to account for some 25–30 percent of the world's publications and citations in the field, with the PS continuing to outperform the Brookhaven AGS in this respect – there was a rapid decline in both the publication and citation totals for the AGS compared with the late 1960s (the citation figure, for example, dropped from 28 percent of the world total in 1966 to 10 percent in 1976). However, in terms of "advances" (papers cited 15 or more times in a year), the CERN PS (9 percent) and ISR (14 percent) were some way behind Fermilab (36 percent) and SLAC (19 percent). Moreover, most of the crucial discoveries which revolutionized the field were made elsewhere, particularly on SPEAR, the storage-ring operated by the Stanford Linear Accelerator Center, SLAC. This highly innovative and relatively cheap accelerator was responsible for no less than 5 out of the 8 papers cited 100 or more times in a year that were published during this period.

---

[15] In November 1973, for example, a neutrino experiment with Gargamelle benefited considerably from the newly increased PS intensity of over 5×10^{12} protons per pulse, an improvement by a factor of 3 over the usual PS intensity [16, p. 369].

[16] Various technical problems were encountered during that year: see, for example, [17, pp. 370–71].

Table 4
Experimental high-energy physics, 1973–1976 [a]

	Numbers of experimental papers [b] published in past two years		Citations [c] to work of past four years		Average citations per paper		Highly cited papers (Number cited n times)			
	1974	1976	1974	1976	1974	1976	$n \geq 15$	$n \geq 30$	$n \geq 50$	$n \geq 100$
CERN ISR (28 + 28 GeV)	40 (3.5%) [d]	60 (5.0%)	830 (12.0%)	710 (9.0%)	14.0	7.4	28 (14.0%)	12 (18.0%)	2 (7.5%)	1 (12.5%)
CERN PS (28 GeV)	230 (20.0%)	280 (23.5%)	1320 (19.5%)	1140 (14.5%)	2.7	2.2	18 (9.0%)	4 (6.0%)	2 (7.5%)	0
DESY (i) 7 GeV (ii) DORIS (4 + 4 GeV)	35 (3.0%)	30 (2.5%)	180 (2.5%)	180 (2.5%)	2.6	2.7 (8.2) [e]	3 (1.5%)	2 (3.0%)	1 (4.0%)	0
Rutherford Nimrod (7 GeV)	35 (3.0%)	20 (1.5%)	120 (2.0%)	60 (0.5%)	2.1	1.1	0	0	0	0
Daresbury NINA (4 GeV)	15 (1.0%)	10 (1.0%)	70 (1.0%)	40 (0.5%)	2.3	1.8	0	0	0	0
Other W. Europe	60 (5.5%)	50 (4.0%)	330 (5.0%)	340 (4.5%)	3.0	3.1	5 (2.5%)	2 (3.0%)	1 (4.0%)	1 (12.5%)
Total W. Europe	410 (35.5%)	445 (38.0%)	2830 (42.0%)	2460 (32.0%)	3.5	2.9	54 (27.5%)	20 (30.0%)	6 (23.0%)	2 (25.0%)
Argonne ZGS (12 GeV)	100 (9.0%)	70 (6.0%)	470 (7.0%)	360 (4.5%)	2.3	2.1	5 (2.5%)	0	0	0
Brookhaven AGS (33 GeV)	155 (13.5%)	95 (8.0%)	760 (11.0%)	780 (10.0%)	2.5	3.2	15 (7.5%)	5 (7.5%)	2 (7.5%)	1 (12.5%)
Fermilab (400 GeV)	105 (9.5%)	175 (15.0%)	780 (11.5%)	1900 (25.0%)	6.9	6.7	71 (36.0%)	26 (39.0%)	9 (34.5%)	0
CEA (i) 6 GeV (ii) BYPASS (3 + 3 GeV)	15 (1.0%)	–	140 (2.0%)	30 (0.5%)	5.9	2.3	2 (1.0%)	2 (3.0%)	1 (4.0%)	0
Cornell (12 GeV)	15 (1.0%)	20 (1.5%)	110 (1.5%)	100 (1.5%)	3.4	3.3	1 (0.5%)	0	0	0
SLAC (i) 20 GeV (ii) SPEAR (4 + 4 GeV)	80 (7.0%)	110 (9.0%)	670 (10.0%)	1330 (17.0%)	4.2	7.1 25.5 [f]	37 (19.0%)	14 (21.0%)	8 (31.0%)	5 (62.5%)
Other US	50 (4.5%)	20 (2.0%)	270 (4.0%)	80 (1.0%)	2.0	1.1	2 (1.0%)	0	0	0
Total US	515 (45.0%)	485 (41.0%)	3190 (47.0%)	4590 (59.5%)	3.3	4.5	133 (67.5%)	47 (70.0%)	20 (77.0%)	6 (75.0%)
Serpukhov (70 GeV)	110 (9.5%)	125 (10.5%)	480 (7.0%)	490 (6.5%)	2.8 (4.6) [g]	2.1 (3.9) [g]	7 (3.5%)	0	0	0
Other – rest of world	115 (10.0%)	125 (10.5%)	260 (4.0%)	200 (2.5%)	1.2	0.8	3 (1.5%)	0	0	0
Total – rest of world	225 (19.5%)	245 (21.0%)	740 (11.0%)	690 (9.0%)	1.9	1.5	10 (5.0%)	0	0	0
World total	1150 (100%)	1180 (100%)	6780 (100%)	7740 (100%)	3.1	3.3	197 (100%)	67 (100%)	26 (100%)	8 (100%)

[a] The notes to tables 7 and 9 of Martin and Irvine [51] also apply.
[b] All publication figures have been rounded to the nearest 5.
[c] All citation figures have been rounded to the nearest 10.
[d] All percentage figures have been rounded to the nearest 0.5%.
[e] This is the figure for DORIS papers only.
[f] This is the figure for SPEAR papers only.
[g] This is the figure for papers in Western journals only.

These papers reported the discoveries of the J/psi (a discovery shared with the Brookhaven AGS), the psi prime, the tau heavy lepton, and the charmed mesons D and D* which were generally seen as representing the first conclusive evidence of explicit charm (there had been earlier hints of charm from Brookhaven and CERN).

To set against this, the CERN accelerators had a number of notable contributions to their credit, though none seems to have had quite the same impact. From the ISR, there was the discovery of the rising total cross-section for proton–proton collisions. However, it was the Serpukhov accelerator which had a few years earlier provided the first indications that hadron–hadron cross-sections were rising (in kaon–proton collisions, at least), although many Western physicists seem not to have been convinced by the result until it was subsequently "confirmed" by the ISR experiment. [17] Consequently, the ISR finding would almost certainly have had a greater impact if it had been the first of its kind rather than the second. Another important, and at the time unique, result from the ISR was the discovery of an anomalously large number of high transverse-momentum events (see footnote 14) which provided indirect evidence for the existence of point-like components ("partons") within hadrons. However, again the first hints of this hadronic structure had been obtained elsewhere (in the deep inelastic scattering experiments at SLAC) which somewhat lessened the impact of the ISR findings. Beside these two discoveries, the ISR was responsible for several, slightly lower-level advances, including the first indications of a "jet" structure in the particles produced in proton collisions, and the discovery of the direct production of "prompt" leptons (electrons and muons) from proton–proton interactions.

In addition, the CERN PS made a number of important advances during this period. One of these arose from a 1973 experiment with the Gargamelle bubble-chamber and demonstrated that neutrino (and antineutrino)–nucleon cross-sections rise linearly with energy. When combined with earlier SLAC work on deep inelastic scattering, these results provided strong experimental evidence that "quarks" were more than just an interesting theoretical concept. In other words, "quarks became real" [29, p. 440].

A second, and far more important, result from the Gargamelle programme on the PS was the discovery of "neutral currents." This provided the first vital experimental evidence in support of the unified theory of electromagnetic and weak interactions [18] put forward by S. Weinberg and A. Salam in 1967. However, perhaps because high-energy physicists had become accustomed to major discoveries being made in the United States rather than Europe, the Gargamelle result was not immediately acknowledged by everyone as constituting conclusive evidence for the existence of neutral currents, with some believing that other interpretations of the observations reported in the paper had not been completely ruled out. [19] Confidence in the discovery was not helped by the fact that initially a strong difference of opinion existed within the Gargamelle collaboration itself over what the experimental data really showed (cf. Hammond [39, p. 372]), some of those who had worked in the past in the field of neutrino physics being amongst the more sceptical. Nor were the CERN management particularly quick in lending their support to the possible discovery. [20] A senior physicist described the situation at CERN in the following terms during an interview:

---

[17] "In 1971, such a growth of the positive kaon–proton total cross-section was reported from Serpukhov but the idea of a cross-section growing again with energy did not, in general, sink in" [18, p. 67]). For a further discussion of the relationship between East European and Western high-energy physics, as well as an evaluation of the scientific performance of the Serpukhov accelerator, see Irvine and Martin [43].

[18] The unification of the four forces of nature (gravitational, electromagnetic, weak, and strong) into a single unified theory has been an elusive dream of physicists since Einstein, and it is the over-riding goal of high-energy physicists today. This first step, to unify two of the four forces, is therefore of vital importance, and has been described as "a synthesis as profound as that achieved by Maxwell when he united the phenomena of electricity and magnetism in one theoretical framework" (Jentschke [48, p. 13]). The authors of the theory have since been rewarded with a Nobel Prize.

[19] Cf. e.g. Barish [5, p. 313]: "It was difficult to make an absolutely convincing case for neutral currents from any single experiment because of the experimental problem that for neutral current reactions *both* the initial and final state neutrinos are non-observables. This means that the reaction is under-constrained and alternate mechanisms for the observations are possible. Rather than actually 'seeing' neutral current events, other explanations for apparent events of this topology must be explained away. For this reason, it was vital to look for neutral currents in a variety of experiments and reactions with different background problems."

> The management at CERN was not confident about the neutral current discovery. Things were definitely held back by the management. They were confused by experiments at Fermilab that first appeared to find neutral currents and then subsequently couldn't find them – they were known as 'alternating' neutral currents for a while! Although this was the major discovery at CERN, many people are still unhappy about the way it was handled. It lost a lot of its value as a result. I talked to friends on the Gargamelle experiment, and there were big divisions inside the collaboration as to whether neutral currents had been established. Those who thought they had, had to argue a lot. If you look at the early announcements of the work in the *CERN Courier*, you will see how hesitant the CERN management was in believing the discovery. Neutral currents were not believed until they were confirmed by Fermilab (Interview, 1981).

These early doubts (which were largely dispelled in 1974 by the publication of a new and more detailed paper on the Gargamelle results) may explain why, despite the significance of the discovery, the main experimental publication reporting the result failed to earn more than 100 citations in a year (although it came quite close), unlike the crucial discoveries made at SLAC and elsewhere over this period. Nevertheless, it remains a discovery of extremely great importance [21], and one which does illustrate the advantages of the methodical approach to high-energy physics adopted at CERN. (We have previously concentrated on the disadvantages.) Taken together, neutral currents and the linear rise of the total neutrino–nuclear cross-section can with some justification be claimed to represent

> well-deserved scientific rewards of very deliberate policies and of efforts pursued over many years. At the PS these policies involved the increase of the PS intensity through the addition of the booster, and the construction of a high quality neutrino beam. The heavy liquid bubble-chamber Gargamelle was built in France and brought to CERN where it was exploited in the neutrino beam (Van Hove and Jacob [67, pp. 67–68]).

One other result from Gargamelle is worth mentioning – the observation of a possible charmed particle announced in March 1975. However, as with several of the other major advances from CERN discussed earlier, this result was not the first of its type to be published. That honour went to a bubble-chamber group working at the Brookhaven AGS accelerator, who, after nine months of attempting to a rule out other interpretations of a candidate event, finally published their result in April 1975. Consequently, the Brookhaven publication seems to have had slightly more impact (at least if judged in terms of numbers of citations), although it too had considerably less impact than the SLAC paper published a year later reporting the discovery of the D meson (since this was generally acknowledged to represent the first *conclusive* evidence for explicit charm).

Despite these notable discoveries from CERN, arguably the most important experimental result [22] of this four-year period, and indeed of the 1970s, was the virtually simultaneous discovery of the J/psi particle on two United States accelerators at Brookhaven (the AGS) and Stanford (SPEAR). In view of the close similarity between the Brookhaven AGS and CERN PS accelerators, there has been much speculation within the high-energy physics community as to why the J/psi was not discovered on the PS (or indeed on the ISR where it was observed very soon after the discovery). In the case of the PS, the reasons were technical as much as organizational. In particular, it should be noted that the J/psi was *only just* detectable at the AGS energy of 30 GeV; because of this, the leader of the research team making the discovery (S. Ting, who is widely acknowledged to be among the best experimental high-energy physicists in the world) had first to go through a very rigorous checking procedure before he was convinced that a new particle had been found (cf. Ting [63]). Undoubtedly it would have been considerably more difficult to make the discovery at 26 GeV, the normal operating energy of the PS accelerator. Hence, in this particular case, the relatively small difference in energy between the two accelerators may well have proved crucial (cf. Van Hove [65, p. 30]) had a similar experiment been attempted on the PS at the same time. Indeed, a second point to note is that a proposal had previously been submitted to carry out on the PS a bispectrometer experiment very similar to that completed on the AGS:

---

[20] Galison [36, p. 499] reports that they were "afraid that CERN would be publicly embarrassed by the forthcoming American announcement" on the failure by the Harvard-Pennsylvania-Wisconsin-Fermilab experiment to detect neutral currents, and that the neutral current discovery might suffer a similar fate to the recently "buried" split A_2.

[21] A number of physicists interviewed argued that the discovery of neutral currents warranted the award of a Nobel Prize. Why it was not thus rewarded is still a matter of speculation.

[22] See footnote 32 in Martin and Irvine [51].

> Unfortunately, its approval was somewhat delayed at CERN as the first version of the proposal was not accepted (*ibid.*)[23]

As was the case with the CP-violation experiment discussed earlier, this missed opportunity seems to have been related in part to the more complex selection procedures then in operation at CERN. At Brookhaven, the approach was comparatively flexible so that an experiment promising important experimental results might still be mounted quickly, even if the formal proposal itself left something to be desired. According to the more critical physicists interviewed by us, the proposed bispectrometer experiment revealed the limitations of the CERN committee procedure system and in particular its tendency, in the past at least, to shy away from taking risks.

As to why the ISR did not discover the J/psi, it should first be noted that it did in fact come very close. One experimental collaboration, which finished collecting data at the ISR in September 1974, two months *before* the discovery was announced, obtained results that were subsequently found to contain five J/psi events (cf. [19, p. 69]). One of those involved in the collaboration, the American, L.M. Lederman, had previously carried out an experiment at Brookhaven that had yielded a somewhat mysterious result (the so-called "shoulder" in the lepton-pair production curve); but instead of profiting from the comparative advantage of the ISR in the experiment to investigate the reason for the anomalous result (which was in fact due to the J/psi), he and the group were side-tracked by the newly discovered phenomena of high transverse-momentum events and prompt leptons (see above). One of the CERN staff members closely associated with the ISR summed up the situation in these terms:

> There was a signal observed here, but people couldn't understand it. I don't know why. Lederman was involved and he is one of the best experimentalists around, but even he didn't see it. The machine was right, and the experiment was right, and even the observation was right. But we still didn't see the J/psi. This is still considered a *tremendous failure* here (Interview, 1981).

We shall, however, leave further analysis of the factors determining the scientific performance of the ISR until later when its outputs up to 1982 have been considered. To sum up, then, it is clear that with a different approach at CERN – in particular, a more flexible decision-procedure on experiments, and a more adventurous attitude on the part of ISR experimentalists – the course of history in experimental high-energy physics between 1973 and 1976 might have been rather different. As it turned out, this was the period when the CERN research facilities lost their relatively short-lived ascendancy over other accelerator centres, and most of the crucial discoveries in these dramatic four years were made by machines on the other side of the Atlantic. In short, the revolution of the mid-1970s which gave rise to the so-called "new physics" was, with the exception of the discovery of neutral currents, largely an all-American affair.

6. 1977 to 1980: CERN consolidates

Probably the main event at CERN between 1977 and 1980 was the completion of the Super Proton Synchrotron (SPS), which began scheduled operation for experimental research on 7 January 1977 after a very rapid and successful commissioning period. This time, unlike with the PS and to a lesser extent the ISR, a full range of sophisticated detectors was ready for immediate experimental use. The four years also saw important developments on the ISR, with the introduction of various second-generation detectors. By now, it had been realized that the early detectors, which had concentrated on small-angle scattering, were inadequate for many important types of experiments, such as the study of "jets" where 4π detectors (covering virtually the complete solid angle) are needed (cf. [21, pp. 394–95]). These developments enabled CERN to enter the late 1970s with the range of sophisticated facilities and equipment needed to take on the task of consolidating the experimental gains made during the previous revolutionary period. In this section, we consider what effects these various developments had on the scientific output from CERN.

---

[23] It was probably this which gave rise to the following comment by a former Director of Brookhaven National Laboratory: "At least we cannot be accused of having disapproved an experiment here which then gave exciting results elsewhere." (Goldhaber [38, p. 82])

Table 5
Experimental high-energy physics, 1977–1980 [a]

	Numbers of experimental papers [b] published in past two years		Citations [c] to work of past four years		Average citations per paper		Highly cited papers (Number cited *n* times)			
	1978	1980	1978	1980	1978	1980	$n \geq 15$	$n \geq 30$	$n \geq 50$	$n \geq 100$
CERN SPS (400 GeV)	25 (2.5%) [d]	80 (8.5%)	320 (4.0%)	520 (8.5%)	12.7	5.0	19 (11.5%)	7 (14.5%)	3 (21.5%)	0
CERN ISR (31 + 31 GeV)	50 (4.5%)	50 (5.5%)	590 (7.0%)	440 (7.5%)	5.4	4.4	11 (6.5%)	2 (4.0%)	0	0
CERN PS (28 GeV)	245 (22.0%)	110 (11.5%)	1170 (14.5%)	770 (12.5%)	2.2	2.2	13 (8.0%)	2 (4.0%)	1 (7.0%)	0
DESY (i) 7 Gev										
(ii) DORIS (5 + 5 GeV)	45	60	450	950	5.7	8.8	36	16	4	0
(iii) PETRA (19 + 19 GeV)	(4.0%)	(6.5%)	(5.5%)	(15.5%)		(11.7%) [e]	(22.0%)	(33.5%)	(28.5%)	
Rutherford Nimrod (7 GeV)	20 (2.0%)	20 (2.5%)	50 (0.5%)	90 (1.5%)	1.2	2.2	1 (0.5%)	0	0	0
Daresbury NINA (4 GeV)	10 (1.0%)	10 (1.5%)	20 –	20 –	1.0	0.7	0	0	0	0
Other W. Europe	35 (3.0%)	20 (2.5%)	220 (2.5%)	130 (2.0%)	2.7	2.3	3 (2.0%)	0	0	0
Total W. Europe	435 (39.0%)	350 (38.0%)	2830 (34.5%)	2920 (48.0%)	3.2	3.7	83 (50.5%)	27 (56.5%)	8 (57.0%)	0
Argonne ZGS (12 GeV)	55 (5.0%)	30 (3.0%)	410 (5.0%)	310 (5.0%)	3.2	3.6	8 (5.0%)	3 (6.5%)	0	0
Brookhaven AGS (33 GeV)	60 (5.5%)	50 (5.5%)	420 (5.0%)	170 (3.0%)	2.7	1.6	0	0	0	0
Fermilab (400 GeV)	180 (16.5%)	180 (19.0%)	2620 (32.0%)	1320 (21.5%)	7.3	3.6	40 (24.5%)	10 (21.0%)	5 (35.5%)	1 (50.0%)
Cornell										
(i) 12 GeV	20	10	80	100	2.3	3.2	4	2	0	0
(ii) CESR (8 + 8 GeV)	(1.5%)	(1.5%)	(1.0%)	(1.5%)			(2.5%)	(4.0%)		
SLAC										
(i) 20 GeV	105	55	1250	690	5.7	4.4	26	6	1	1
(ii) SPEAR	(9.5%)	(6.0%)	(15.0%)	(11.5%)	(12.5) [f]		(16.0%)	(12.5%)	(7.0%)	(50.0%)
Other US	15 (1.0%)	5 (0.5%)	80 (1.0%)	30 (0.5%)	2.3	1.4	0	0	0	0
Total US	435 (39.0%)	330 (35.5%)	4850 (59.0%)	2620 (43.0%)	5.3	3.4	78 (47.5%)	21 (43.5%)	6 (43.0%)	2 (100%)
Serpukhov (70 GeV)	135 (12.0%)	130 (14.0%)	320 (4.0%)	310 (5.0%)	1.2	1.2	0	0	0	0
Other – rest of world	115 (10.0%)	120 (13.0%)	190 (2.5%)	240 (4.0%)	0.8	1.0	3 (2.0%)	0	0	0
Total – rest of world	250 (22.5%)	250 (27.0%)	510 (6.0%)	550 (9.0%)	1.0	1.1	3 (2.0%)	0	0	0
World total	1115 (100%)	930 (100%)	8190 (100%)	[illegible] (100%)	3.5	3.0	164 (100%)	48 (100%)	14 (100%)	2 (100%)

[a] The notes to tables 7 and 9 of Martin and Irvine [51] also apply.

[b] All publication figures have been rounded to the nearest 5.

[c] All citation figures have been rounded to the nearest 10.

[d] All percentage figures have been rounded to the nearest 0.5%.

[e] This is the figure for PETRA papers only.

[f] This is the figure for SPEAR papers only.

The figures in table 5 reveal that the three CERN accelerators together accounted for between 25 and 30 percent of the world output of experimental publications in the period, earning a similar proportion of the world total of citations. In terms of individual accelerators, the contributions of the ISR and PS declined somewhat from the levels achieved previously, but nowhere near as dramatically as the Brookhaven AGS (it earned only 3 percent of citations in 1980 compared with 23 percent ten years earlier) which virtually dropped out of contention as a front-line accelerator as staff at the centre concentrated their efforts on the construction of the new collider, ISABELLE. [24] To some extent, the declines of the PS and ISR were compensated for at CERN by the growing output and impact of the SPS which, taking advantage of beams and detectors that represented a significant improvement over those at Fermilab, achieved a particularly high rate of citations per paper. (The figure of 12.7 citations per paper in 1978 was similar to that for SPEAR at Stanford in 1978 and for PETRA at DESY in 1980.) Nevertheless, of all the major accelerators, the Fermilab machine seems to have contributed most during this period, with a large publication output, over one quarter of total world citations, and a relatively high figure for citations per paper. It should be noted, however, that the record of the Fermilab machine did decline markedly in the latter part of the period, not only because of competition from the SPS, but also because of budgetary pressures arising from an ambitious and expensive capital-expenditure programme (on an "Energy Doubler/Saver" extension to the accelerator and a new proton–antiproton collider facility) that severely curtailed the laboratory's experimental activity.

A similar picture is evident when one considers "advances" (papers cited 15 or more times in a year). Here, the CERN accelerators together actually achieved more "advances" than those at any other laboratory – 43 compared with 40 at Fermilab, 36 at DESY, and 26 at SLAC. However, in terms of this indicator, the Fermilab machine was the most successful individual accelerator. For advances of a slightly higher level (papers cited 30 or more times in a year), the DESY colliders (PETRA and DORIS) appear to have done best with 16 such papers compared with 11 for the CERN machines and 10 for Fermilab. At CERN, the main contributions from the ISR included the work on single leptons, and various observations of charmed-particle production. However, as is evident from table 5, the ISR was rather overshadowed by the achievements of the SPS: these included the refutation of the "high-y anomaly" observed at Fermilab (see Paper I [51]); measurements of structure-functions and deviations from scaling which were found to be in agreement with the predictions of quantum chromodynamics (currently regarded as the best candidate theory of strong interactions); observations of prompt neutrinos in "beam dump" experiments which were believed to provide evidence for charmed-particle production and possibly a third type of neutrino (the tau neutrino); and confirmation of the di-muon events first seen at Fermilab, together with a vast amount of new data (dwarfing previous world statistics) on this form of particle interaction sufficient to enable the explanation to be cleared up. (Until then, various competing hypotheses had been proposed, but the data were not adequate to decide between them.) Analysis of these advances made by the SPS shows that a significant part of the accelerator's research programme was essentially concerned with repeating earlier, sometimes crude, "first-generation" Fermilab experiments, either to correct some of the mistakes and remove various anomalies, or to confirm Fermilab findings but with superior statistics. While a certain degree of satisfaction was inevitably derived by CERN experimentalists from the former [25], the latter task of confirming previous results is inevitably seen by most researchers as being much less rewarding than making the initial discovery, however much better the statistics produced in the confirming experiment. As the leader of one of the main neutrino experiments at the CERN SPS observed,

> I would gladly trade them our events for their discovery (J. Steinberger, quoted in Walgate [68, p. 279]).

The citation records of papers reporting discoveries suggest that these generally have more impact on the advance of scientific knowledge than even the highest-quality confirmatory experiment. This probably explains why, in terms of highly cited papers, the Fermilab accelerator continued to have a better overall record in this respect than the CERN SPS. However, these figures do not reveal the full story about the relative contributions made by the Fermilab and SPS machines. In particular, it should be noted that most of the highly cited papers from Fermilab were published in 1977 (seven out of the ten papers cited more than 30 times, for example), and that a number of them (three out of the five cited more than 50 times) reported results (on trimuons) that subsequently came under severe doubt (see Paper I [51]). Hence, on balance, the SPS was probably not as far behind the Fermilab accelerator in terms of major advances as the figures in table 5 might first suggest, in particular after 1978.

---

[24] Cf. [23, p. 260]: "The considerable effort being devoted to the ISABELLE project has necessarily weakened the support of the 33 GeV AGS and its experimental programme."

[25] For example, a report on the 1977 international conference at Budapest described how "CERN ... was eager to spread the word that the 'high-y anomaly' and related abnormalities in high energy neutrino interactions had been wiped out by new [SPS] data" [24, p. 226].

What cannot be doubted, however, is that, as in previous periods, crucial discoveries continued to elude the CERN machines. There were two such discoveries made during these four years, the upsilon at Fermilab, and parity violation in inelastic electron-scattering at the Stanford Linear Accelerator Center – the discovery which, in the eyes of many physicists, finally removed any lingering doubts about the Weinberg–Salam theory being *the* unified theory of electro-weak interactions.[26] At one stage, it did in fact seem as though the SPS had made a discovery of equally crucial importance – a meson containing the bottom quark. This result, if it had subsequently been confirmed, would have constituted a discovery of similar magnitude to that of the charmed [illegible] mesons found a few years earlier at Stanford: at the time, it was heralded as providing "seemingly conclusive evidence for the fifth quark" (Robinson [57, p. 777]). Unfortunately, the "particle" later disappeared under a mass of new data (cf. [31, p. 241]),[27] thus denying the SPS one of its best chances of a significant discovery.

As in the case of other discoveries, physicists have speculated why the upsilon was discovered at Fermilab and not at CERN. Probably the SPS came into operation somewhat too late, and certainly at the time of the Fermilab announcement it was not in a position to mount an appropriate experiment.[28] In contrast, the upsilon *could* almost certainly have been discovered on the ISR, and, as with the J/psi three years earlier, ISR users came very close. During a preliminary experimental run on the ISR in 1976 (a year before the Fermilab discovery), a few events were seen at an energy of 9.5 GeV, which is almost exactly the mass that the upsilon was subsequently found to have. However, when one of the experimenters involved reported the result,

> he was greeted with great scepticism, so he went away, and eventually came back with new data without this signal. The collaboration only 'found' the signal again after the Fermilab discovery. So, as with the J/psi, the upsilon was in fact seen here first, but it was not interpreted (Interview, 1981).

It is worth noting that, if these few early events had been observed, say, at Fermilab, they would in all probability have been publicly announced, just as in early 1976 a Fermilab collaboration reported on the basis of just 11 events, the discovery of a new particle at 6 GeV (which they named the "upsilon" – see [22, p. 83]), another particle which then "disappeared" in subsequent months. At CERN, in contrast, the tradition has generally been to adopt a more cautious approach to announcing potentially interesting research findings. While this means that CERN has, at least since the "split A_2," largely avoided "mistaken" results of the type that initially afflicted Fermilab (one exception being the candidate "bottom meson" referred to above), it may also have resulted in the laboratory yielding "priority" in a few cases to experiments carried out elsewhere. This question of the difference in research style between Europe and the United States, with high-energy physicists in the latter tending to adopt a more bold and speculative, but at the same time more risky, approach to their experiments, is one to which we

---

[26] Although neutral currents had been discovered in 1973, this Stanford experiment was the first to demonstrate that they had just the properties predicted by the Weinberg–Salam theory.

[27] More recently, in 1981, another CERN experiment, this time on the ISR, claimed to find a "bottom particle." However, this finding was disputed, in particular by another ISR team (see [55, p. 478] for details), and most high-energy physicists seem to have reserved their judgement as to whether to believe the result.

[28] "At CERN, the energies available from the SPS for the counter experiments which have appropriate detector configurations in the West Area are too low for fruitful upsilon hunting." [25, p. 320]

Table 6
Experimental high-energy physics, post-1980 [a]

	Numbers of experimental papers published			Citations [b] to work of past four years			Average citations per paper			Highly cited papers (Number cited n times)	
	1980	1981	1982	1980	1981	1982	1980	1981	1982	$n \geq 15$	$n \geq 30$
CERN $p\bar{p}$ (270 + 270 GeV)	–	4 (1.0%) [c]	8 (2.0%)	–	–	110 (2.0%)	–	–	9.3	3 (20.0%)	3 (75.0%)
CERN SPS (400 GeV)	42 (9.5%)	72 (16.5%)	58 (15.0%)	520 (8.5%)	610 (12.0%)	800 (16.0%)	5.0	3.6	3.8	2 (13.5%)	0
CERN ISR (31 + 31 GeV)	19 (4.5%)	26 (6.0%)	36 (9.5%)	440 (7.5%)	420 (8.0%)	530 (10.5%)	4.4	3.8	4.7	1 (6.5%)	0
CERN PS (28 GeV)	45 (10.5%)	41 (9.5%)	26 (7.0%)	770 (12.5%)	430 (8.5%)	250 (5.0%)	2.2	1.6	1.4	0	0
DESY PETRA (20 + 20 GeV)	25 (5.5%)	23 (5.5%)	31 (8.0%)	470 (7.5%)	580 (11.0%)	880 (17.5%)	11.7	9.2	9.3	6 (40.0%)	0
DESY DORIS (5 + 5 GeV)	6 (1.5%)	7 (1.5%)	3 (1.0%)	460 (7.5%)	280 (5.5%)	120 (2.5%)	10.2	7.3	4.4	0	0
Other W. Europe	23 (5.5%)	17 (4.0%)	11 (3.0%)	250 (4.0%)	160 (3.0%)	130 (2.5%)	1.8	1.4	1.5	0	0
Total W. Europe	160 (36.5%)	190 (44.0%)	173 (44.5%)	2920 (48.0%)	2480 (48.0%)	2820 (55.5%)	3.7	3.3	3.9	12 (80.0%)	3 (75.0%)
Brookhaven AGS (33 GeV)	32 (7.5%)	20 (4.5%)	11 (3.0%)	170 (3.0%)	170 (3.0%)	120 (2.5%)	1.6	1.7	1.4	0	0
Cornell CESR (8 + 8 GeV)	4 (1.0%)	5 (1.0%)	12 (3.0%)	50 (1.0%)	160 (3.0%)	200 (4.0%)	–	18.2	9.3	2 (13.5%)	0

Fermilab (400 GeV)	72 (16.5%)	66 (15.5%)	50 (13.0%)	1320 (21.5%)	1060 (20.5%)	800 (16.0%)	3.6	3.1	2.7	0	0
SLAC PEP (18 + 18 GeV)	–	–	10 (2.5%)	–	–	20 (0.5%)	–	–	2.4	0	0
SLAC SPEAR (4 + 4 GeV)	12 (2.5%)	9 (2.0%)	16 (4.0%)	360 (6.0%)	270 (5.0%)	300 (6.0%)	6.4	5.5	6.6	1 (6.5%)	1 (25.0%)
SLAC linear accelerator (32 GeV)	13 (3.0%)	7 (1.5%)	8 (2.0%)	330 (5.5%)	240 (4.5%)	130 (2.5%)	3.2	3.0	2.7	0	0
Other US	19 (4.5%)	17 (4.0%)	10 (2.5%)	390 (6.5%)	260 (5.0%)	140 (2.5%)	3.0	2.8	2.0	0	0
Total US	152 (34.5%)	124 (29.0%)	117 (30.0%)	2620 (43.0%)	2160 (41.5%)	1710 (33.5%)	3.4	3.2	3.0	3 (20.0%)	1 25.0%)
Serpukhov (70 GeV)	64 (14.5%)	65 (15.0%)	45 (11.5%)	310 (5.0%)	350 (6.5%)	300 (6.0%)	1.2	1.4	1.3	0	0
Other – rest of world	62 (14.0%)	52 (12.0%)	54 (14.0%)	240 (4.0%)	210 (4.0%)	250 (5.0%)	1.0	0.9	1.1	0	0
Total – rest of world	126 (29.0%)	117 (27.0%)	99 (25.5%)	550 9.0%)	560 (10.5%)	550 (11.0%)	1.1	1.1	1.2	0	0
World total	438 (100%)	431 (100%)	389 (100%)	6090 (100%)	5190 (100%)	5070 (100%)	3.0	2.7	2.9	15 (100%)	4 (100%)

---

[a] The notes to tables 7 and 9 of Martin and Irvine [51] also apply

[b] All citation figures have been rounded to the nearest 10

[c] All percentage figures have been rounded to the nearest 0.5%

will return in our discussion below on the factors affecting the scientific performance of each of the CERN accelerators. It will suffice to say here that CERN has in recent years become more aggressive in promoting the results of its research, and this, together with the concerted action taken to develop a unique range of sophisticated detectors and specialized particle beams, has played a large part in enabling the users of the laboratory to match and eventually surpass the performance of American laboratories in recent years, as we shall see in the next section.

7. Experimental high-energy physics, post-1980: The European renaissance

In June 1980, the CERN SPS was turned off for 11 months to make possible the construction of the new proton-antiproton collider. As one commentator remarked at the time,

> such a major sacrifice of prime research time has never been seen before and is as sure a pointer as any to the importance of the extension in CERN's research potential which the proton–antiproton collider will bring [30, p. 143]

In the event, this major sacrifice very quickly began to pay dividends. The first proton–antiproton collisions were observed in July 1981, but during the remainder of 1981 the luminosity remained low – a peak of 5×10^{27} cm^{-2}s^{-1} was achieved, a factor of 200 below the design figure. By the end of the following year, however, the luminosity had been increased by an order of magnitude and this, combined with an almost continuous period of operation of the collider lasting nearly two months, yielded a vastly improved integrated luminosity (of some 25 inverse nanobarns). At the end of this run, it was noted that,

> According to the theory, this should be sufficient for the production of some intermediate bosons of the charged kind [33, p. 7]

It therefore came as little of a surprise when, a few weeks later in January 1983, it was announced that the first signs of the charged W intermediate vector boson had been seen. This was followed in May 1983 by the announcement of the discovery of the neutral Z_0 boson.

Besides the commissioning of the proton–antiproton collider, CERN users also benefited from several other improvements to the research facilities. For example, while the SPS was shut down, the WA experimental area for the SPS was upgraded, while the new European Hybrid Spectrometer was completed in 1981, giving SPS users a wide range of powerful detectors, including the EMC, CHARM, BEBC, and Omega prime facilities, with which to exploit the accelerator. In the case of the ISR, improvements continued to be made to the luminosity, and the first experiments involving collisions between alpha particles and between protons and antiprotons were carried out in 1980 and 1981 respectively.

The effects of these various additions and improvements to the experimental facilities at CERN are readily apparent in the figures in table 6. To take first the output of experimental publications, although the number yielded by the SPS dropped slightly in 1982 (reflecting the shut-down a year earlier), this was more than compensated by the emergence of the first results from the proton-antiproton ($p\bar{p}$) collider, and by an appreciable increase in ISR papers as users of the machine rushed to complete as many experiments as possible before it closed at the end of 1983 (see Paper III [52]). Consequently, whereas papers from the CERN accelerators accounted for 24.5 percent of the world total in 1980, the corresponding figure in 1981 and 1982 was some 33 percent. Similarly, in terms of citations, CERN users increased their fraction of the world total from 28.5 percent in 1980 to 33.5 percent two years later. There were three reasons for this: first, the coming into operation of the collider; second, the growing impact of work on the SPS – SPS papers virtually doubled their share of world citations from 8.5 percent in 1980 to 16 percent in 1982, finally catching up Fermilab (although in the meantime the SPS had been overtaken by PETRA at DESY), and, third, a significant increase in the overall impact of ISR papers from 7.5 percent of the world total of citations in 1980 to 10.5 percent in 1982.

The figures on citations per paper suggest that, in 1981 and 1982, CESR at Cornell and PETRA at DESY were the machines with the highest average impact per paper, both recording a level of 9.3 in 1982. For the CERN SPS, the corresponding figure fell from 5.0 in 1980 to 3.6 in 1981 (although it did rise again very slightly in 1982), comparatively low for an accelerator that had only been operating for five years. This somewhat rapid obsolescence can perhaps best be explained by the fact that, by 1980, the Fermilab accelerator – which had been operational since 1972 – and the SPS had between them carried out most of the important work in the 400 GeV energy-range. If so, this illustrates the dangers of building a machine very similar in energy to one that has already been operating for several years, even if it does represent a significant technical improvement on that earlier accelerator. The figures on citations per paper in the case of the ISR are particularly illuminating in this respect since they show an increase from 3.8 in 1981 to 4.7 in 1982. The latter is remarkably high for a machine that has been operating for twelve years (far higher, for example, than the very successful Brookhaven AGS in its twelfth year – see table 3), and must surely be linked to the fact that the ISR was until its closure in 1983 always a unique facility. Given that no larger proton–proton collider is likely to be completed in the near future, the decision to close the ISR in order to free resources for the construction of LEP clearly represented a major sacrifice for the facility's users. In contrast, publications from the CERN PS earned only 1.4 citations per paper – very similar to the figures for the Brookhaven AGS and Serpukhov – suggesting that it no longer constituted a frontier research facility in high-energy physics.

The last two columns of table 6 give the numbers of papers published in 1981 and 1982 that had been highly cited by the end of 1982. These figures suggest that, in terms of "advances" (papers cited 15 or more times in a year), PETRA has proved the most successful over this period with six such papers, followed by the CERN proton–antiproton collider with three, the SPS and CESR with two each, and the ISR with one (Fermilab had no such papers). If the figures for CERN and DESY are combined, one sees that Western Europe produced 80 percent of the papers cited 15 or more times, while the United States managed only 20 percent. A similar picture emerges in the case of "major advances" (papers cited 30 or more times in a year) – the CERN collider yielded three such papers, while the United States had just one (from SPEAR), this being the first time ever that a CERN machine has led the world in terms of the most highly cited papers. Overall, the figures in table 6 for the various machines at CERN and DESY give good grounds for claiming that the onset of the 1980s had witnessed a European renaissance in experimental high-energy physics, even before the discoveries of the W and Z particles were announced in 1983.

Table 7
Comparison of the scientific outputs of the CERN PS and Brookhaven AGS accelerators, 1961–1980

		1961–1964		1965–1968		1969–1972		1973–1976		1977–1980	
		PS	AGS	PS	AGS	PS	AGS	PS	AGS	PS	AGS
Experimental papers [a]		21.5%	8.0%	28.5%	19.5%	25.0%	16.5%	22.0%	10.5%	17.5%	5.5%
Citations to recent [b] work [a]		14.5%	23.0%	27.5%	27.5%	22.0%	20.5%	17.0%	10.5%	13.5%	4.0%
Average citations per paper		1.9	8.0	3.3	5.5	2.5	3.2	2.5	2.8	2.2	2.2
Number of papers cited *n* times [a]	$n \geq 15$	10.0%	27.0%	30.5%	28.0%	15.5%	18.0%	9.0%	7.5%	8.0%	0.0%
		(8.5%) [c]	(39.0%)	(27.5%)	(33.0%)	(13.0%)	(17.5%)	(13.0%)	(7.5%)	(7.0%)	(0.0%)
No. of times achieved	$n \geq 30$	6.0%	29.5%	5.5%	66.5%	3.0%	5.0%	6.0%	7.5%	4.0%	0.0%
		(7.0%)	(55.0%)	(13.0%)	(66.5%)	(4.5%)	(2.5%)	(7.0%)	(11.5%)	(4.0%)	(0.0%)
	$n \geq 50$	0.0%	25.0%	0.0%	100%)	0.0%	0.0%	7.5%	7.5%	7.0%	0.0%
		(0.0%)	(57.0%)	(0.0%)	(100%)	(0.0%)	(0.0%)	(9.0%)	(11.0%)	(4.5%)	(0.0%)
	$n \geq 100$	0.0%	100%)	–	–	0.0%	0.0%	0.0%	12.5%	0.0%	0.0%
		(0.0%)	(100%)			(0.0%)	(0.0%)	(0.0%)	(19.0%)	(0.0%)	(0.0%)

[a] All figures are expressed as % of the relevant world total, and each has been rounded to the nearest 0.5%.

[b] Work published during the preceding 4 years.

[c] The figures in brackets correspond to those for the number of times papers earned *n* citations in one year.

8. The overall performance of the CERN accelerators

8.1. The Proton Synchrotron

Having analyzed in some detail the historical development of experimental high-energy physics over the period from 1961 to the end of 1982, we are now in a position to draw some more general conclusions about the overall scientific performance of each of the principal CERN accelerators. Table 7 summarizes the data on the outputs from the first of these accelerators, the Proton Synchrotron, relative to those from the rival AGS machine at Brookhaven. As can be seen, the PS produced appreciably more experimental papers than the AGS. However, this is the only respect in which it consistently outperformed the AGS, since the average impact of those papers was in general considerably less. In terms of citations per paper, the Brookhaven machine until very recently had a superior record to the PS. The AGS was also by far the more successful of the two accelerators in terms of both "crucial discoveries" (papers cited 100 times or more in a year) and "major advances" (those cited 50 or 30 times or more), particularly during the 1960s. Certainly, in the case of "advances" (papers cited 15 or more times), the PS overtook the AGS in the mid-1970s, but it had been well behind in the early years. Hence, when taken together, these indicators strongly suggest that, overall, the AGS was appreciably more successful than the PS over the twenty-year period under consideration. As noted at the time of its twentieth anniversary, the AGS

> has been one of the most productive of accelerators for physics, counting amongst its discoveries the muon neutrino, charge-parity violation, the omega minus, the J/psi,

Table 8
Assessments (on a 10-point scale [a]) of main proton accelerators in terms of (1) discoveries, (2) providing more precise measurements

	Self-rankings	Peer-rankings	Overall rankings
(1) Discoveries			
Brookhaven AGS	9.5(±0.1) ($n = 48$)[b]	9.0(±0.1) ($n = 121$)	9.2(±0.1) ($n = 169$)
CERN PS	7.1(±0.2) ($n = 86$)	6.7(±0.2) ($n = 83$)	6.9(±0.1) ($n = 169$)
CERN ISR	6.8(±0.3) ($n = 36$)	5.9(±0.2) ($n = 133$)	6.1(±0.2) ($n = 169$)
CERN SPS	5.9(±0.3) ($n = 68$)	5.6(±0.2) ($n = 101$)	5.7(±0.1) ($n = 169$)
Fermilab 400 GeV	7.4(±0.3) ($n = 46$)	7.1(±0.1) ($n = 123$)	7.2(±0.1) ($n = 169$)
Serpukhov	3.8(±0.5) ($n = 20$)	2.6(±0.1) ($n = 149$)	2.7(±0.1) ($n = 169$)
(2) More precise measurements			
Brookhaven AGS	7.1(±0.2) ($n = 47$)	7.2(±0.2) ($n = 120$)	7.2(±0.1) ($n = 167$)
CERN PS	8.5(±0.1) ($n = 86$)	8.5(±0.1) ($n = 81$)	8.5(±0.1) ($n = 167$)
CERN ISR	7.3(±0.3) ($n = 35$)	6.9(±0.2) ($n = 132$)	7.0(±0.1) ($n = 167$)
CERN SPS	8.2(±0.2) ($n = 67$)	8.2(±0.2) ($n = 100$)	8.2(±0.1) ($n = 167$)
Fermilab 400 GeV	6.3(±0.2) ($n = 46$)	6.0(±0.2) ($n = 121$)	6.1(±0.1) ($n = 167$)
Serpukhov	4.3(±0.5) ($n = 20$)	3.5(±0.2) ($n = 147$)	3.6(±0.2) ($n = 167$)

[a] 10 = top, and 1 = bottom. The rankings are based on the relative outputs from the accelerators over their entire careers up to the time of the interviews with the high-energy physicists in late 1981/early 1982. The figures in brackets indicate the root-mean-square variations, giving some approximate idea of the divergence of opinion within each group. It should be noted that these rankings were normally made by interviewees in written form.

[b] The sample size of each group is denoted by n.

the charmed baryon, and numerous other particles and resonances [32, p. 242].

Indeed, the AGS can boast a record of major discoveries that only the Stanford Positron–Electron Accelerator Ring, SPEAR, can match. The PS, in contrast, could manage only one result of similar importance – the discovery of neutral currents. This difference between the two accelerators suggested by the indicators listed in table 7 is also well supported by the results obtained in interviews with 182 high-energy physicists (see Paper I [51] for details). Interviewees were invited to assess the overall scientific performance of the world's six principal proton accelerators on a 10-point scale (10 = top, 1 = bottom) in terms of two different criteria: (1) crucial experiments and discoveries; and (2) experiments involving more precise measurements of known particles and their properties but without discovering anything new. The results are given in table 8. On this 10-point scale of achievement, the AGS was judged considerably ahead of the PS in the case of the first criterion (9.2 compared with 6.9), a view, it should be stressed, held by users of both accelerators.

A rather different conclusion, however, emerged when the second criterion of "precise measurements" was considered – the PS was rated 8.5 on the 10-point scale, somewhat ahead of the AGS on 7.2 (with the figures based on self-evaluation and peer-evaluation again showing little variance). Two of the bibliometric indicators are of relevance to this result – the total number of publications, and the average citations per paper. As can be seen from table 7 above, the AGS had a rather better record in terms of the latter. However, this advantage appears to have been outweighed by the far greater publication output of the PS, particu-

Table 9
Factors explaining the relative scientific performance of the CERN PS and Brookhaven AGS (% of interviewees believing this factor to have been important [a])

		Users of AGS (%) ($n = 42$)	Users of PS (%) ($n = 69$)	Users of AGS and PS (%) ($n = 17$)	All interviewees (%) ($n = 125$)
1.	More bold, speculative ethos of US physicists – Europeans more conservative, less risky approach	55	54	65	51
2.	Greater experience of US physicists – Europeans had to learn to use a large accelerator	31	42	47	36
3.	US had better experimentalists in 1960s	29	29	29	30
4.	Higher work ethic and more competitive attitude of US physicists	29	20	35	20
5.	Europe's "missing generation" of physicists – wartime emigration to US of many of the best	19	14	29	11
6.	Luck of Brookhaven in choosing right experiments	21	13	24	17
7.	Better scientific management at Brookhaven – e.g. quick to respond to the unexpected	21	14	24	16
8.	CERN's tendency to "over-engineer" – e.g. detectors built bigger but later	10	14	12	10
9.	Problems of CERN being multinational – e.g. slow committees, over-conservative choice of experiments	62	46	59	51
10.	Social structure of European groups more hierarchical – non-scientific ("political") factors introduced	17	4	6	10

[a] These figures represent minimum values only since they are based on a content analysis of answers to a general question concerning the factors structuring the relative scientific performance of the two accelerators. The sample size of each group of interviewees is denoted by n.

larly in the 1970s when the "citations per paper" figure was not very different for the two accelerators. As a result, whereas between 1961 and 1964 AGS experiments appear to have had a significantly greater impact (at least in terms of the total numbers of citations), the two machines had a very similar record in the period from 1965 to 1968, and then the PS edged ahead. The gap widened considerably during the 1970s as the output from the AGS declined, citations to work published over the preceding four years falling from 20 percent of the world total in 1969–72 to 4 percent in 1977–80. Thus, over their entire lifetimes, the record of the PS was probably slightly superior in terms of this indicator, a result consistent with the rankings given in table 8 for experiments involving more precise measurements.

Besides identifying differences between the scientific performance of the PS and AGS, high-energy physicists were also asked to point to reasons for those differences. Even though this question was unstructured, in the sense that interviewees were not prompted as to all possible reasons for the divergence in performance, the responses tended to conform to certain patterns, and it was therefore relatively easy to classify them into a number of separate categories (although the factors are obviously related to a certain extent). The results are summarized in table 9, which lists all factors cited by 10 percent or more of the interviewees. The first five of these ten factors relate to differing characteristics of the respective user-communities of the AGS and PS. Particularly important here was what many interviewees (in both Europe and North America) identified as a more bold, speculative ethos of physicists in the United States, where the research system was regarded as normally allowing more scope for individual initiative. In contrast, physicists in Europe were seen as adopting a more solid, safe, and conservative approach to their experiments. As can be seen from the table, this factor was specified by approximately half those questioned on the issue, and by rather more (65 percent) of those who had direct experience of using both the AGS and PS accelerators.

Another extremely important factor appears to have been the generally higher level of experimental experience among the AGS user-community, particularly in the 1960s. Many Brookhaven users had previously carried out experiments on the laboratory's 3 GeV Cosmotron or the 6 GeV Berkeley Bevatron, while the experience of most European experimentalists was limited to accelerators of 1 GeV energy or less. The jump involved in moving from experiments on the 0.6 GeV CERN synchro-cyclotron (or small national accelerators) to the 28 GeV PS was obviously much more difficult than for experimentalists progressing from the relatively high-energy 6 GeV Bevatron to the 33 GeV AGS. Consequently, users of the PS inevitably took rather longer to establish how best to exploit the accelerator than their counterparts on the AGS. This factor was again particularly stressed by those who had used both accelerators (47 percent compared with 36 percent of all interviewees).

Other interviewees (30 percent) were rather less specific about the nature of the differences between PS and AGS users, merely mentioning that the US had more competent experimentalists in the 1960s, a fact which some attributed to a better postgraduate educational system. Also cited were the higher work ethic and more competitive attitude of American researchers (20 percent), and Europe's "missing generation" of physicists lost through both the Second World War and emigration to the United States in the 1930s and 1940s (11 percent).

Besides these five user-related explanations, a sixth factor identified was the "luck" of Brookhaven in choosing the right experiments (17 percent). Some of those specifying this factor did, however, go on to relate this "luck" to the greater flexibility and even opportunism of the Brookhaven management in re-orienting the laboratory's experimental programme whenever an unexpected new possibility presented itself (16 percent).[29] Both this and factor 8 in table 9 (CERN's tendency to "over-engineer"[30] – for example, constructing highly sophisticated detectors that came into operation some time later than less advanced instruments built by their competitors) imply at least some criticism of managerial decision-making at CERN. Indeed, certain high-energy physicists evidently believe that the quality of the Brookhaven management was in no small way responsible for the position of advantage achieved by the AGS over the PS. However, these two factors also need to be seen in the context of factor 9 relating to CERN's multinational character. As a central laboratory catering for a large number of geographically dispersed users, CERN inevitably was under great pressure to provide a service that was, above all, reliable, and this meant that those facilities

---

[29] See also the quotation above in the section on 1961–64 about the absence of committees in the early years at Brookhaven.

[30] This tendency was attributed by some interviewees to the fact that CERN has a relatively large number of "in-house" technical staff who build a higher proportion of experimental equipment than is the case at US accelerators, where universities have a greater role in providing equipment. One consequence of this, it was argued, is that American researchers have generally been more in control of their experiments, and, because they understood the equipment better, were able to undertake rapid modifications when the need arose.

> must be designed and constructed *conservatively* and within, if at the limit of, current industrial technology or at least with all innovations studied and tested beforehand (Hine [41, p. 180], emphasis added).

As a previous Director-General has stated, CERN has tried to

> avoid taking risks with the reliability of the machine because then all its users suffer and the worst thing of all is to launch accelerator projects irrespective of whether or not one knows how to overcome the technical problems.[31] That is the surest way of ending up with an expensive machine of doubtful reliability, later than was promised, and a physicist community which is thoroughly dissatisfied. CERN, I am glad to say, has avoided this trap ...[32] (Adams [1, p. 23]).

In general, however, the physicists interviewed pointed to the disadvantages associated with CERN's multinational character, rather than to the advantages. In particular, the need to balance national interests has resulted, it was argued, in a generally greater reliance on a sometimes cumbersome committee structure than at the US national laboratories. Decision-making, therefore, has often been slower, the response to new initiatives sometimes over-cautious, and the choice of experiments too conservative, at least in the opinion of over 50 percent of those interviewed, although many of these did point to the decision by CERN to proceed with the proton–antiproton collider as evidence of a marked improvement in this respect.

Another, wider institutional problem (factor 10 in table 9) centres on the social structure of European experimental collaborations. Not only have these tended to be slightly larger than their American equivalents; they have also apparently been seen by many physicists as more hierarchical and more subject to influence by non-scientific considerations. For example, in Europe, for various historical reasons, an academic culture has evolved in which there may have been a greater unwillingness among young researchers to risk offending senior professors than was typical in the United States. This has sometimes manifested itself in a reluctance to challenge conventional wisdom, and in the adoption of a risk-minimizing approach to experiments. This was contrasted by several interviewees with the more aggressive and risky approach adopted by many young Americans (see factor 1 in table 9), an approach which may fail, but sometimes succeeeds dramatically. One senior experimentalist, who has worked on both sides of the Atlantic, described the situation thus:

> The institutional structures of physics in the US and Europe are different. The people making discoveries at the AGS – they were from universities, and their character was formed in a particular social situation. To get a tenured job in a good American university, you have to do well recognized work and to make an individual contribution. In a European country, things are different. The approach is one of career-optimization. It leads to researchers in Europe working on 'precise measurements'-type experiments, rather than trying to challenge existing ideas. Professors are much stronger – you have to fit in with their established programmes of work (Interview, 1981).

To sum up, most of the factors believed by high-energy physicists to account for the differences between the scientific performance of the PS and the AGS extend far beyond the perimeters of the CERN laboratory itself, relating either to its user-community or to the wider institutional context of the centre and of research in Western Europe. Only factors 7 and 8 relate directly to

---

[31] The author probably had in mind here the experiences of Brookhaven in the late 1970s with the ISABELLE project. When the project began, the laboratory was by no means certain that the large number of superconducting magnets required could be manufactured. Unfortunately, the gamble did not pay off, with the result that there were considerable delays and escalating costs – see, for example, Broad [6], and Paper III [52]. The project was finally terminated in 1983.

[32] Whether CERN in fact knew how to overcome all the likely technical problems when it embarked on the ISR project is open to dispute. At the time the ISR began operation, its construction was described as "doing something entirely new whose successful operation *could not be guaranteed*" ([14, p. 31], emphasis added). Similar reservations might also apply to the more recent proton–antiproton collider project, where it was by no means certain that the design luminosity could be reached.

CERN policy and management, and these were by no means the most frequently cited factors. This point is worth stressing lest the identification of differences between the performance of the PS and AGS accelerators be seen as attributing blame unduly to CERN itself: the experience with the PS must rather be seen as at least partly reflecting certain broader features of European society during this period.

8.2. The Intersecting Storage Rings and the Super Proton Synchrotron

Assessing the performance of the ISR and SPS is nowhere near as straightforward a task as for the PS, where a direct comparison could be drawn with the AGS because of the similarity in energy and in the starting date of the first experiments. In contrast, the ISR was always a unique facility, while the SPS, although identical in energy to the Fermilab accelerator, only began scheduled operation for experimental research in 1977, more than four years after the American machine. (This was, as we have noted earlier, a rather large lead to concede to Fermilab, given that much of the physics in a new energy region tends to be carried out within the first few years of becoming accessible.) Nevertheless, there are grounds for believing that valid comparisons between these and other machines can still be drawn. For example, we can contrast the CERN ISR with the Serpukhov 70 GeV accelerator in view of the fact that they began producing experimental publications within two years of each other, and both were the highest-energy machines of their type in the world. And, in the case of the SPS, besides comparing its outputs directly with those from Fermilab, it is also possible to examine the scientific performance of Fermilab before and after the SPS started operating to determine what effect this new accelerator had.

Table 10 summarizes for the fourteen-year period from 1969 to 1982 the overall scientific outputs of these four large proton machines and the subsequent impact their work created within the scientific community. For the first four years, Serpukhov publications gained a relatively high rate of citations per paper (CPP), especially for articles in Western journals (for the four years, the figure was 7.7, having been as high as 10.7 in 1970 for the first two years of the accelerator's operation), suggesting these early results had a considerable impact. The same is true of the ISR, with a corresponding figure of 12.9 over the first two years of its experimental life. However, because the Serpukhov accelerator had been operating longer, it achieved more citations overall up to 1972 – 4 percent of the world total in that year compared with 2.5 percent for the ISR. In terms of "advances" and "major advances," the records of the two accelerators were virtually identical. Serpukhov was, however, responsible for the sole "crucial discovery" (cited more than 100 times in a year) of this period – the discovery of rising total cross-sections for hadron–hadron collisions. Finally, it should be noted that the new machine at Fermilab began to make an impact on the statistics for highly cited papers towards the end of 1972.

Over the next four years from 1973 onwards, the picture presented by the bibliometric data in the table is one of clear Fermilab dominance, except in the case of very highly cited papers (gaining 100 or more citations in a year) where the ISR did better (as a result of the publication reporting the discovery of rising total cross-sections for proton–proton collisions). Both Western machines were by then contributing significantly more than the Serpukhov synchrotron, which was increasingly overshadowed by the much higher-energy facility at Fermilab.

In the four years up to 1980, Fermilab continued to produce a large number of experimental publications (17.5 percent of total world output), and achieved by far the greatest number of citations and highly cited papers of the four major proton machines. However, some of the mistaken results from earlier experiments (see Paper I [51]) were beginning to be recognized, and these should be taken into account when comparing its scientific performance with that of the SPS. As we noted earlier, it is significant that the SPS figure for citations per paper fell very rapidly from 12.7 in 1978 to under 4 in 1981 and 1982. Clearly, by the time the SPS began operating in 1977, most if not all the more obvious experiments to be carried out in the energy region had been attempted at Fermilab, and the major discoveries (like dimuons and the upsilon) had already been made. Thus, while the SPS was undoubtedly responsible for a number of major advances (cited 30 or 50 times in a year) stemming from sophisticated second-gener-

Table 10
Comparison of the scientific outputs of the Serpukhov, CERN ISR, Fermilab (FL) and CERN SPS accelerators, 1969–1982

		1969–1972			1973–1976			1977–1980				1981–1982	
		Serpukhov	ISR	FL	Serpukhov	ISR	FL	Serpukhov	ISR	FL	SPS	FL	SPS
Experimental papers [a]		4.5%	1.0%	0.0% [c]	10.0%	4.0%	12.0%	13.0%	5.0%	17.5%	5.0%	14.0%	16.0%
Citations to recent [b] work [a]		4.0%	2.5%	0.0% [c]	7.0%	10.5%	18.5%	4.5%	7.5%	27.5%	6.0%	18.0%	14.0%
Average citations per paper		3.4 (7.7) [d]	12.9	–	2.4 (4.2) [d]	9.9	6.8	1.2 (2.4) [d]	4.9	5.5	6.4	2.9	3.8
Number of papers cited n times [a]	$n \geq 15$	9.5% (12.5%) [e]	10.5% (11.5%)	3.5% (4.0%)	3.5% (3.0%)	14.0% (15.0%)	36.0% (32.0%)	0.0% (0.0%)	6.5% (7.0%)	24.5% (23.0%)	11.5% (13.0%)	0.0% (0.0%)	13.5% (12.0%)
	$n \geq 30$	19.0% (23.0%)	19.0% (18.5%)	14.5%) (14.0%)	0.0% (0.0%)	18.0% (15.5%)	39.0% (34.0%)	0.0% (0.0%)	4.0% (4.0%)	21.0% (24.5%)	14.5% (13.0%)	0.0% (0.0%)	0.0% (0.0%)
	$n \geq 50$	22.0% (26.5%)	22.0% (13.5%)	33.5% (33.5%)	0.0% (0.0%)	7.5% (11.0%)	34.5% (23.5%)	0.0% (0.0%)	0.0% (0.0%)	35.5% (36.5%)	21.5% (18.0%)	–	–
	$n > 100$	100.0% (100.0%)	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)	12.5% (6.5%)	0.0% (6.0%)	0.0% (0.0%)	0.0% (0.0%)	50.0% (66.5%)	0.0% (0.0%)	–	–

[a] All figures are expressed as a % of the relevant world total, and each has been rounded to the nearest 0.5%.

[b] Work published during the preceding 4 years.

[c] The very first experimental papers from Fermilab were published in the latter part of 1972, but the statistics are too small to be significant, except those for highly cited papers.

[d] This is the figure for papers in Western journals only.

[e] The figures in brackets correspond to those for the number of times papers earned n citations in one year.

ation experiments, its record up to 1982 was probably not quite as distinguished as that of the ISR in its early years. In particular, it should be noted that, while the ISR was responsible for one crucial discovery (cited more than 100 times in a year), the SPS did not come close to equalling this feat in the period reviewed.

In summary, then, if these four accelerators are compared in terms of "discoveries," the Fermilab machine must undoubtedly be placed first, followed by the ISR, and with little to choose between the aggregate records of Serpukhov and the SPS. The fact that the SPS achieved in the four years up to 1980 approximately the same number of papers cited more than 15, 30 and 50 times in a year as Serpukhov achieved over its first twelve years (Serpukhov has contributed very little since 1972), would probably tip the balance in favour of the SPS being ranked third, slightly ahead of Serpukhov. Referring back to the peer-evaluation results given earlier in table 8, one sees that these rankings are in close agreement with those made by high-energy physicists: Fermilab was given a score of 7.2 on the 10-point scale for discoveries (though, it should be noted, some way behind the Brookhaven AGS), while the ISR was scored 6.1, the SPS 5.7, and Serpukhov 2.7.[33]

As for the relative positions of the four machines in terms of "precise measurements," the results are rather more difficult to interpret, largely because of the differing periods of time over which the accelerators have been operating. However, we have seen from tables 5 and 6 that since 1980 the SPS has yielded more publications than the three other machines, and that from 1978 SPS publications have earned more citations per paper (except in the case of ISR publications in 1981 and 1982). It is therefore perhaps not too surprising that, for "precise measurement" work, high-energy physicists ranked the SPS first of these four accelerators (although it was placed a little behind the PS) with a score of 8.2, while the ISR was also rated quite highly at 7.0 in line with its consistently high

Table 11

Factors explaining the relative scientific performance of the CERN ISR (% of interviewees believing this factor to have been important[a])

		Users of ISR (%) ($n = 29$)	Others (%) ($n = 84$)	All interviewees (%) ($n = 113$)
1.	Poor/wrong choice of detectors initially (e.g. no 4π solid-angle detectors)	72	49	55
2.	ISR given too low priority by CERN (because of SPS) – some experiments/detectors too late	34	19	23
3.	Poor scientific management – failure to respond promptly when need for new detectors apparent	28	19	21
4.	Wrong choice of machine – colliding beams less fruitful for HEP than fixed target machines	3	23	18
5.	Research programme too fragmented – no overall strategy (e.g. too many small experiments initially)	28	13	17
6.	Early luminosity too low for successful experiments	34	17	21
7.	Problems with the main detector (Split Field Magnet)	7	11	10
8.	Had to learn from scratch how to use machine – proton–proton collisions inherently very complex	48	40	42
9.	Bad luck – turned out not to be a very exciting energy region	17	31	27

[a] Minimum estimates only (see note a to table 9). The sample size of each group of interviewees is denoted by n.

---

[33] These results in all likelihood exaggerate the difference between the contributions made by the Soviet accelerator and the other three. Part of the difference may be attributable to an ignorance on the part of many Western scientists of all the results from the Soviet machine – see Irvine and Martin [43].

BEST COPY AVAILABLE

number of citations per paper. Fermilab, despite its large number of papers, was deliberately marked down by many interviewees because of several earlier mistaken results, scoring 6.1, with Serpukhov again ranked last on 3.6.

Having examined the relative outputs from the CERN ISR and SPS, we can now move on to analyse the factors that have determined their scientific performance. Taking the ISR first, there can be no doubt that, from a technical point of view, the machine has been a tremendous success. In particular, its luminosity has been improved by several orders of magnitude over its lifetime, far surpassing the original design specifications. There is thus some justification to the claim that

> The ISR is widely regarded as the most perfect example to date of the accelerator builder's art [27, p. 231].

However, there is, as has been seen, a widespread feeling among high-energy physicists that, as a research facility, the ISR was for some years not exploited as successfully as it might have been, somehow missing the opportunity of making such crucial discoveries as the J/psi and the upsilon. This sense of disappointment that the potential of the machine was not fully realized (which is not to say that it was unsuccessful) is reflected in table 11. This summarizes the results obtained from asking high-energy physicists to identify the main factors determining the ISR's relative scientific performance. Factors 1 to 5 relate to problems associated with the management of this facility and to decisions taken at CERN, these clearly being seen as central by many of those interviewed. In particular, the decisions to concentrate initially on small-angle collisions came in for much criticism, especially from ISR users (with 72 percent citing this as a major problem). One of the CERN senior staff closely connected with the ISR recalled the events surrounding this decision in the following terms:

> The Split Field Magnet is the main detector on the ISR. The design of the device as originally instrumented emphasized forward and backward angles, not large angles. So it missed the J/psi exactly. At the time, CERN management was convinced that no particles would be found at large angles. This was based on Argonne experiments and cosmic rays. There was some competition between different possible spectrometers, but they chose the wrong one. At one stage, they considered doing two detectors, one for looking at large angles. But they decided to phase these, and leave the large-angle one till later. They thought the ISR had the energy region to itself. They forgot about the competitive aspect, and Ting [the co-discoverer of the J/psi] at Brookhaven. (Interview, 1981)

The decision to postpone construction of a large-angle detector was also probably related to factor 2 in the table, the judgement by some high-energy physicists (23 percent of these interviewed) that the ISR and its experimental programme were not given sufficient priority by CERN management. The most critical of those physicists believed that the ISR was regarded by some senior CERN officials as little more than a stop-gap measure to preserve the laboratory's accelerator-building capacity (and capital-equipment budget) until agreement could be secured to build the more prestigious SPS accelerator. It was pointed out that, in order to facilitate approval for the SPS project in 1971, CERN had to promise to reduce the budget allocated to existing facilities (in Laboratory I, as it was then called) by just over 200 million Swiss francs between 1971 and 1978 (see Paper I [51]). In 1973, the Director-General reported that preparations for the SPS were

> beginning to dominate all our planning at CERN since it must be pushed hard during the next three years in order to start the experiments as early as possible. This makes considerable manpower and financial demands on Laboratory I. (Jentschke [48, p. 23]).

While most of this heavy burden fell on the PS experimental programme, the ISR did not escape entirely.

Some of the high-energy physicists interviewed (21 percent) were not so specific in their criticisms, citing the scientific management of the ISR in general as a major problem (factor 3). Others (17 percent) argued that, at least in the early years, the ISR experimental programme had been too fragmented, there being no overall strategy for exploiting the accelerator and addressing the outstanding theoretical questions of the day (factor 5). As one physicist commented,

> They tried to use the ISR in the same way as the PS, with lots of small experiments. (Interview, 1981)

Initially, there was, for example, no large general-purpose detector of the type that is nowadays regarded as essential to exploit colliders. In addition, the early experiments were rather too crowded together – there were at one stage four experiments at just one of the intersection regions. Such problems were perhaps compounded by the division of opinion that existed (and still exists) within

BEST COPY AVAILABLE

Table 12
Factors explaining the relative scientific performance of Fermilab and the CERN SPS (% of interviewees believing this factor to have been important [a])

	Users of Fermilab (%) (*n* = 39)	Users of SPS (%) (*n* = 56)	Users of Fermilab and SPS (%) (*n* = 7) [b]	All interviewees (%) (*n* = 136)
1. Fermilab had 4-year lead – SPS too late	56	54	29	58
2. Fermilab more bold, speculative experiments – CERN solid, precise 2nd-generation experiments	49	36	43	39
3. Problems of CERN being multinational – e.g. slow committees, over-conservative choice of experiments	10	9	14	11
4. More resources and technical support for CERN experiments – sometimes inadequate at Fermilab	49	55	43	50
5. SPS a much better accelerator – e.g. better beams	15	32	29	26
6. Fermilab cut too many corners building accelerator and detectors – unreliable, "under-designed"	31	32	29	31
7. Fermilab philosophy wrong – too much of running costs channelled into new accelerator projects	15	13	29	14
8. Fermilab spread themselves too thinly on the ground with experiments, given the resources	21	14	14	13
9. Energy range of both accelerators turned out to be relatively unexciting	13	5	0	10

[a] Minimum estimate only (see note a to table 9). The sample size of each group of interviewees is denoted by *n*.

[b] This is a very small sample size, and the % figures may not therefore be statistically very significant

the laboratory over the value of the ISR as an experimental facility. A not insignificant fraction of physicists believed the ISR to have been the wrong choice of machine (factor 4), arguing that colliding protons is not as fruitful a research technique as the "fixed target" approach with synchrotrons, and that the ISR diverted CERN's efforts away from the much more important SPS project.

Problems with the management of the ISR were by no means the only factors cited. Many physicists also believed that technical problems were central in limiting the accelerator's scientific performance, particularly in the early years: 21 percent of interviewees pointed out that the luminosity of the ISR was initially too low for certain types of experiment (factor 6); and 10 percent specified various technical difficulties with the main detector. In addition, two problems (factors 8 and 9) were identified that were probably completely unavoidable. Nearly half those interviewed argued that, because the ISR was the world's first proton collider, physicists had to start from scratch in establishing how to use it,[34] while the very big jump in centre-of-mass energy that it made possible meant that there was necessarily a long learning-curve in finding how best to exploit the facility. Moreover, proton–proton collisions are inherently difficult to interpret – they involve the interaction of three quarks with three other quarks, generally producing large numbers of secondary particles which make analysis of the results a rather complex process.[35] In addition, a certain number of interviewees (27 percent) put forward the view that the ISR had been the victim of bad luck, in the sense that the energy range which it opened up proved not to be as exciting as had been anticipated. However, this seems difficult to reconcile with the fact that the J/psi and the upsilon – the two most highly cited discoveries of the 1970s, and arguably the two with the greatest impact – were there to be unearthed by the ISR if appropriate experiments had been carried out soon enough.

---

[34] In contrast, users of the Stanford electron–positron ring, SPEAR, were able to draw upon the experience gained from earlier, smaller rings at Frascati, Orsay, and Cambridge.

[35] The analysis of the data requires a powerful computing capability, and some of those interviewed pointed out that this was not available until several years after the ISR began operating.

Let us now move on to consider the factors underlying the scientific performance of the SPS, especially in relation to its Fermilab rival. As we have seen, the Fermilab accelerator had up to 1982 proved somewhat more successful in terms of major advances and discoveries, while the SPS had a better record (according to the peer-evaluation results) for precise measurements. Table 12 summarizes the main reasons given by high-energy physicists to explain these differences in the pattern of performance of the two accelerators. Factors 1 and 2, and perhaps also 3, together largely explain Fermilab's better record from major advances. Well over half of those interviewed (58 percent) felt that the Fermilab lead of some four and a half years proved too big a handicap to the SPS (factor 1). To many, the subsequent difficulties for the SPS could have been predicted; as early as 1973, it had been pointed out that such a lead represents

> nearly half the 'interesting' lifetime of most high energy physics installations – which raises the question of whether there will be significant investigations left for it to pursue. (Hammond [40, p. 1120])

Indeed, it was the recognition by CERN in 1976 of the significance of Fermilab's lead that led them to consider, even at this early stage, other possibilities for exploiting the SPS. It was as a response to this that the proton–antiproton collider proposal was put forward and approved relatively quickly (cf. Van Hove [66, p. 31]).

In addition, this extensive lead had implications for the type of experimental programme that could profitably be mounted on the SPS accelerator. Many of the physicists interviewed pointed to a difference in approach between Fermilab and CERN, with the former undertaking more bold, speculative experiments (see factor 2). This is, in turn, partly a function of the different styles of American and European researchers (a factor discussed earlier in connection with the AGS and PS accelerators), and partly the result of deliberate policy choices made by the two laboratories. At Fermilab, they decided to take full advantage of the lead over the SPS by carrying out a broad, but sometimes rather shallow, experimental programme, trying a little of everything in order to "cream off" any major discoveries that could be made in the energy range. The discovery of dimuons and the upsilon were the rewards of this approach, while conversely the production of several papers containing "mistaken" results was a negative consequence of the strategy. At CERN, in contrast, it was recognized early on that the SPS would stand the best chance of making a significant contribution to scientific progress by performing detailed, high-precision, second-generation experiments. Though this often meant repeating previous Fermilab work, these experiments have tended to provide definitive sets of statistics, as well a clearing up the mistakes made by the American accelerator. Factor 3 – the problems associated with the multinational character of CERN – has already been considered in connection with the PS, and will not be discussed further. It will suffice to record that 11 percent of interviewees regarded this as a limiting factor on the performance of the SPS.

Factors 4 to 8 shed further light on the reasons why the SPS has done rather better at "precise measurement" experiments (a subject already touched upon in our treatment of factor 2, above). Undoubtedly the main reason for this is the greater resources and technical support available at CERN, this factor (4) being mentioned by about half the researchers interviewed.[36] Many of the research staff at Fermilab openly expressed envy of the large, sophisticated detectors then available at CERN; there was no way, they argued, that Fermilab, with its limited operating budget (which averaged just under half that of CERN[37]) could afford to build such technically advanced equipment. (In the early 1980s, Fermilab experienced difficulty even in meeting the power costs of running its experimental programme, and was forced to shut down the accelerator for extended periods.)

Also of great importance has been the high technical quality of the CERN SPS: with the Fermilab machine so far ahead, there was little point in rushing to complete the construction programme of the facility. The SPS was therefore built very carefully, and is, as a result, a significantly better accelerator. Thirty-two percent of SPS users and 15 percent of Fermilab users identified this as a factor explaining the SPS's advantage in producing precise results. The obverse of this is that the Fermilab accelerator has proved somewhat unreliable, over 30 percent of physicists at both CERN and Fermilab pointing to this problem (factor 6). The magnets in particular have caused numerous problems, initially absorbing moisture from the humid air with the result that over half had to be replaced, some more than once. This generally lower reliability at the American centre can be understood partly as a result of long-term financial pressures on the laboratory,[38] and partly as a consequence of the philosophy of R.R. Wilson, the first Director of Fermilab – a philosophy which involved

---

[36] In addition, 84 percent of the physicists completing the attitude survey (described in the final section of this paper) agreed with the statement, "CERN provides European high-energy physicists with a better level of technical support and facilities than that enjoyed by Americans at their national laboratories". A mere 2 percent disagreed.

[37] See table 5 in Paper I [51].

> cutting corners whenever possible and generally following a tight design. This approach is given credit for getting the accelerator built quickly, and within a stringent budget .... But some physicists now question whether a more conservative approach – such as that being followed [with the SPS at CERN] .... would really have required any more time or money (Hammond [40, p. 1117]).[39]

Another aspect of Wilson's philosophy was his belief that a significant proportion of Fermilab funds should be invested in accelerator development to ensure that, once the SPS came into operation, the Fermilab facility could be rapidly upgraded to give it a renewed advantage over CERN. Some of those interviewed (about 14 percent – see factor 7) clearly felt that too high a proportion of Fermilab's operating budget had been channelled into the new Energy Doubler/Saver project,[40] thus impoverishing what was already a rather inadequately supported experimental programme. A not insignificant number of interviewees (13 percent, and 21 percent at Fermilab) also felt that, in its effort to "skim the cream," Fermilab had tried to do too many experiments too quickly.[41] By the end of the first four years of operation, 152 experiments had been completed, well over twice as many as at the SPS (66 experiments) over the equivalent period of time. This clearly seems to have had implications for the experimental results obtained by Fermilab users. As one commentator concluded,

> The laboratory had started life with a shotgun approach to particle research, mounting many small experiments, and critics think Wilson waited too long to consolidate the experimental program and to build fewer large selected experiments with greater resolving power (Metz [53, pp. 196–97]).

Finally, we should note that 10 percent of interviewees expressed a feeling that, in retrospect, neither accelerator had made a very major contribution to high-energy physics because the energy range that they covered turned out to be relatively unexciting. This could not, of course, have been predicted in advance, but it does highlight the negative consequence, identified by certain physicists, of both Europe and the United States making their largest investment of the 1970s in essentially identical accelerators.

This concludes our analysis of the factors determining the relative scientific performance of the three main CERN accelerators. The only remaining task of this paper is to synthesize the principal points arising from our assessment of CERN over the past twenty years, in particular drawing out any lessons of relevance to the subject of Paper III [52], the future prospects for CERN.

9. CERN's past performance – an overall assessment

Before summarizing the main conclusions that can be drawn from our assessment of CERN's past scientific performance, it perhaps needs re-emphasizing that certain aspects of CERN's activities have *not* been evaluated here. We have not, for example, commented upon the very substantial

---

[38] Cf. [28, p.111].

[39] Nearly twice as many of those interviewed agreed as disagreed (61 percent compared with 33 percent) with the statement, "The early Fermilab philosophy of cutting all possible corners to save time and money has not paid off in the long run."

[40] Cf. Metz [53, p. 196]. The Doubler/Saver is a major project to increase the energy of the Fermilab accelerator from 400 GeV to between 500 and 1000 GeV. Paper III [52] gives further details.

[41] In the attitude survey (see final section), 72 percent of those interviewed agreed with the statement, "At Fermilab, there has been a tendency in the past to accept too many experiments, with the result that experiments have often been rushed or prematurely cut short", only 8 percent disagreed.

technical contributions to high-energy physics for which CERN has been responsible. These include the invention of proportional wire chambers, the development of streamer chambers (following up the pioneering work of G. Chikovani at Tbilisi in the Soviet Union), liquid argon detectors, various bubble-chamber developments (ultrasonic chambers, small high-resolution chambers, holographic chambers, etc.), and the invention of the technique of stochastic cooling. Nor have we evaluated the extensive contributions made by CERN theorists – to the quark model, to Regge and Cabibbo theory, to quantum chromodynamics, and more recently to grand unification theories and supersymmetry. And we have mentioned only in passing the participation of CERN physicists in experiments at Serpukhov, and in particular their role in helping make the crucial discovery there of rising total cross-sections.

Furthermore, it should be noted that CERN has been responsible for several wider contributions besides helping to further our knowledge of high-energy physics. These have been described in the following terms:

> It has been 'a source of European spirit'. It has played a key role in re-establishing the stature of European science. It has a continuing impact on science teaching in universities. It has promoted and helped sustain technical excellence in scientific equipment [20, p. 262].

To take the first of these, there was no doubt in the minds of the high-energy physicists we interviewed that CERN has substantially stimulated international co-operation in science, not only in Europe, but also at a world level, promoting contacts with North America, Eastern Europe, and the Third World. As for the second, a cursory glance at tables 1 and 6 to contrast the respective scientific outputs of CERN in the early 1960s and twenty years later is sufficient to demonstrate the tremendous strides made by West European high-energy physicists in relation to their American counterparts. West European experimental papers for example, earned only 19.5 percent of the world citation total in 1964, a factor of four less than the US figure of 77.5 percent. In 1982, the corresponding figures were 55.5 percent and 33.5 percent, showing that Europe had completely reversed the situation. The third effect, the impact on university education, is harder to gauge; undoubtedly, there has been some impact, but whether this is greater than would have been the case if the resources invested in CERN had been spent on other areas of scientific research is impossible to judge. A similar reservation applies to the fourth type of contribution – technological "spin-off." Again, many instances of spin-off have clearly arisen, particularly to firms supplying equipment to CERN, as Schmied [60] has amply documented. However, the "opportunity costs" have also to be taken fully into consideration – is the level of technological spin-off higher than it would have been if the resources spent on CERN had instead been used to support some other type of research, such as exploration of the ocean bed, for example? [42] To this question, there are as yet no ready answers. [43]

This said, it should nevertheless be stressed that we do not feel our assessment is intrinsically weakened by the fact that it has focused almost exclusively on

> CERN's main purpose [which] is to provide Europe's scientists with excellent facilities for high-energy physics research. *All* justification of the investment that is called for from the twelve Member States begins with a belief in the value of such research ([20, p. 262], emphasis added).

Our aim has been to produce systematic, reliable and reproducible conclusions on the extent to which significant contributions to the advance of scientific knowledge have been made by experimental high-energy physicists using the CERN research facilities.

Three main sets of conclusions can be framed on the basis of the assessment outlined above and in Paper I [51]. The first is that, since about 1970 (and with the possible exception of a few years in the mid-1970s when the Fermilab results were having a major impact), the overall record of the CERN machines taken together has been better than that of the accelerators at any other laboratory in the world when judged in terms of experiments producing precise measurements and results with high statistics. Evidence for this comes not only from the figures on total citations (which give some indication of the overall impact of published work), but also from the peer-evaluation results presented in Paper I [51]. Prior to 1970, this position was held by Brookhaven, and before that (from the mid-1950s until the early 1960s) by Berkeley (with the Bevatron).

---

[42] For further discussion of this question, see Irvine and Martin [44].

[43] It is, however, noteworthy that in the attitude survey (described below), nearly twice as many physicists disagreed with the following statement as agreed with it: "The level of resources spent on high-energy physics can be completely justified by the technological spin-off it generates."

The second major conclusion is that, in terms of scientific productivity – that is, scientific performance evaluated in relation to inputs (number of users, funding, etc.) – the record of each of the three US National Laboratories, and the Stanford Linear Accelerator Center in particular, seems to have been significantly better than that of CERN. Supporting evidence for such a conclusion is provided by the figures in tables 8, 10 and 14 in Paper I [51]. It is also noteworthy that, in an attitude survey [44] conducted among the high-energy physicists interviewed in the study, the great majority (71 percent) agreed on the whole with the statement that, "Overall, the American national laboratories have been more cost-effective than CERN in providing experimental high-energy physics facilities," while only 18 percent disagreed.

The third principal conclusion is that, with the exception of neutral currents, the crucial discoveries in high-energy physics between 1961 and 1982 were all made at laboratories other than CERN. As the Director for International Co-operation at the German Federal Ministry for Research and Technology is reported as saying in 1980,

> CERN has been better at building superb accelerators than at discovering spectacular physics (Dr. G. Lehr, reported in Walgate [69, p. 706]).

This feature of CERN's performance was clearly recognized by high-energy physicists whom we interviewed in 1981 and 1982. Of these, 72 percent agreed with the statement, "The CERN accelerators have been responsible for a relatively small number of major discoveries compared with the other main high-energy physics centres," well over three times the number who disagreed (22 percent). Even at CERN, this relative failure did not go unnoticed. In 1980, the Research Director-General admitted that,

> In the last few years, the most important developments in hadron spectroscopy have been those going beyond SU(3) symmetry and connected with the 4th and 5th quarks. Here the role of CERN has been very modest as compared with the discovery of the J/psi at Brookhaven and at SLAC in 1974, the discovery of the first charmed mesons at SLAC in 1976, and the discovery of the upsilon at Fermilab in 1977 (Van Hove and Jacob [67, p. 33]).

These, it should be stressed, are generally regarded as the three most important discoveries of the 1970s, the papers reporting these findings being the only ones in the decade to earn 150 or more citations in a year. Up to 1982, no CERN publication had come close to equalling these figures or having an equivalent impact on the advance of knowledge. However, the situation may have since changed radically with the publication in 1983 of the papers reporting the first observations of the W and the Z particles – likely to be seen as two of the most important discoveries of the 1980s.

During the 1960s, the comparative failure of the Proton Synchrotron to make major discoveries at a time when several were being made on the Brookhaven AGS accelerator can, as we have seen, be explained largely in terms of factors over which CERN itself had little influence – in particular, the relative lack of experience of PS users, the time inevitably taken by a large multinational organization in evolving efficient managerial structures and procedures, and the traditionally hierarchical structure of European research activity. However, in the case of the Intersecting Storage Rings, which clearly *could* have been first to discover the J/psi and the upsilon, several of the main factors identified by physicists as having been responsible are linked to the scientific management of this facility and to decisions taken at CERN. Happily, it would seem that many of the lessons from the experiences with the ISR have been absorbed at CERN, evidence for such a learning process having taken place coming from the very rapid and hugely successful exploitation of the proton–antiproton collider. As for the relative lack of success of the Super Proton Synchrotron in making major discoveries, the crucial factor was the four-year lead held by Fermilab. This cannot be regarded as entirely the fault of European politicians. If the original design for the accelerator produced by CERN had been less expensive, then the American lead would probably have been far shorter.

---

[44] The attitude survey consisted of approximately 30 statements relating to issues previously discussed in the interview. Interviewees had to circle a number on a 7-point scale depending on whether they agreed strongly (1), agreed (2), agreed but with reservations (3), were neutral or had mixed views (4), and so on up to (7), disagreed strongly with the statement.

Perhaps the crucial policy question raised by the above analysis is the extent to which the factors that have limited the scientific performance of the CERN accelerators, in particular those which are internal to the CERN laboratory itself, are intrinsic to any large-scale multinational venture of this sort. Is it inevitable that a laboratory where decisions tend to be made by formal committees rather than individuals or small informal advisory groups, and which must be accountable to a dozen political masters, will take less gambles and risks than a laboratory free from such contstraints? Most high-energy physicists interviewed did seem to feel that, "The system by which the allocation of time on CERN accelerators is decided tends to encourage too much routine research rather than highly innovative but risky experiments" (51 percent agreed with this statement in the attitude survey, appreciably more than the 31 percent who disagreed). The question is whether an international laboratory can devise procedures to overcome this conservative tendency towards safe but routine research.

While we shall not attempt to answer this question here, it is necessary to note two points. First, the decision in 1978 to proceed with construction of the proton–antiproton collider did represent a major gamble by CERN, showing that the conservative tendency can be successfully resisted. Second, it would seem that high-energy physics in the United States is now becoming subject to political pressures similar to those exerted on CERN. In the past, the situation in the large US laboratories was such that individuals like M. Goldhaber at Brookhaven, W.K.H. Panofsky at Stanford, and R.R. Wilson at Fermilab, could stamp their authority on the direction in which their laboratories moved. (Directors-General at CERN were limited by the fact that they were appointed for five years only and, except in one case, were not reappointed.) However, there has been a gradual contraction in the number of US high-energy physics laboratories during the 1970s, so that there are now just three major accelerator centres and a single smaller one (and even this may be more than the US will be able to afford in coming years [45]). This means that an increasingly large number of scientific and institutional interests have to be accommodated (and represented) within the decision-making structure at each centre. Unless the United States proves more adept in avoiding the dangers of what many high-energy physicists termed "committee science" than CERN was in the past,[46] then the previous advantage of US physicists over the rest of the world in terms of making most of the crucial discoveries may disappear permanently. To a certain extent, the balance of power in experimental high-energy physics between North America and Western Europe seems, therefore, to have reached a turning point. A full analysis of the future prospects for CERN, and of whether there are indeed such grounds for European optimism in the longer term forms the subject of Paper III [52].

References

[1] J.B. Adams, The Development of CERN, 1970 to 1980, *CERN Annual Report 1980* (CERN, Geneva, 1980) pp. 17–25.

[2] J.B. Adams, Reflections on a Big Science, *Science and Public Policy* 9 (1982) 81–87.

[3] N.V. Baggett, *AGS 20th Anniversary Celebration* (Brookhaven National Laboratory, BNL 51377, 1980).

[4] J.H. Bannier, Foreword, *CERN Annual Report 1965* (CERN, Geneva, 1965) p. 3.

[5] B.C. Barish, Experimental Aspects of High Energy Neutrino Physics, *Physics Reports* 39 (1978) 279–360.

[6] W.J. Broad, Magnet Failures Imperil New Accelerator, *Science* 210 (1980) 875–78.

[7] *Brookhaven National Laboratory Annual Report July 1, 1970* (Brookhaven, Upton, Long Island, New York, 1970).

[8] C.C. Butler, High Energy Nuclear Physics: Achievements and Future Developments, *Nature* 206 (1965) 763–69.

[9] *CERN Annual Report 1964* (CERN, Geneva, 1964).

[10] *CERN Annual Report 1969* (CERN, Geneva, 1969).

[11] *CERN Annual Report 1972* (CERN, Geneva, 1972).

[12] Some Physics Post-Lund, *CERN Courier* 9 (1969) 232–35.

[13] 1959–1969: Ten Years in the Life of a Machine, *CERN Courier* 9 (1969) 337–46.

[14] Comment, *CERN Courier* 11 (1971) 31.

[15] XVI International Conference on High Energy Physics, *CERN Courier* 12 (1972) 315–17.

[16] PS Intensity Record Broken, *CERN Courier* 13 (1973) 369–70.

[17] Experiments with BEBC, *CERN Courier* 13 (1973) 370–71.

[18] The Proton at Very High Energies, *CERN Courier* 13 (1973) 67–68.

[19] CERN News, *CERN Courier* 15 (1975) 67–70.

[20] Paying Our Way? An Analysis of the 'Economic Utility' of CERN, *CERN Courier* 15 (1975) 262–64.

---

[45] See the pessimistic discussion of the financial future for US high-energy physics in Trilling [64].

[46] The ultimate example of the dangers of "committee science" is, however, perhaps to be found in the performance of East European research centres – see Irvine and Martin [43].

[21] ISR Workshop, *CERN Courier* 16 (1976) 394–95.

[22] And Another?, *CERN Courier* 16 (1976) 83.

[23] Brookhaven – Work on ISABELLE, *CERN Courier* 16 (1976) 258–60.

[24] The Discovery of the Upsilon, *CERN Courier* 17 (1977) 223–27.

[25] Upsilon Hunting, *CERN Courier* 17 (1977) 319–21.

[26] Projects Galore at Brookhaven, *CERN Courier* 18 (1978) 247–51.

[27] A Brief History of CERN, *CERN Courier* 19 (1979) 228–32.

[28] Fermilab: Operation Successes, *CERN Courier* 19 (1979) 111–12.

[29] Looking Back Over the 1970s, *CERN Courier* 19 (1980) 439–41.

[30] Proton–Antiproton Colliding Beams Coming Nearer, *CERN Courier* 20 (1980) 143–51.

[31] Madison Physics Conference, *CERN Courier* 20 (1980) 238–42.

[32] Brookhaven Looks Back … and Forward, *CERN Courier* 20 (1980) 242–45.

[33] When Antimatter Mattered, *CERN Courier* 23 (1983) 6–7.

[34] G. Cocconi, Research at the PS: Electronics Experiments, *CERN Courier* 9 (1969) 349–51.

[35] V. Fitch, A Discovery in Inner Mongolia, in Baggett [3, 49–54].

[36] P. Galison, How the First Neutral-Current Experiments Ended, *Reviews of Modern Physics* 55 (1983) 477–509.

[37] H. Georgi, A Unified Theory of Elementary Particles and Forces, *Scientific American* 244 (No. 4, 1981) 40–55.

[38] M. Goldhaber, AGS 20th Anniversary Celebration, in Baggett [3, 81–84].

[39] A.L. Hammond, Neutral Currents: New Hope for a Unified Field Theory, *Science* 182 (1973) 372–74.

[40] A.L. Hammond, The Big Accelerators: A Progress Report, *Science* 182 (1973) 1117–20.

[41] M.G.N. Hine, Industrial Implications of Large, International Research Enterprises, *CERN Courier* 8 (1968) 179–83.

[42] G. 't Hooft, Gauge Theories of the Forces Between Elementary Particles, *Scientific American* 242 (No. 6, 1980) 90–116.

[43] J. Irvine and B.R. Martin, Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators, *Social Studies of Science* (forthcoming).

[44] J. Irvine and B.R. Martin, What Direction for Basic Scientific Research?, in M. Gibbons, P. Gummett and B.M. Udgaonkar (eds.), *Science and Technology in the 1980s and Beyond* (Longman, London, 1984).

[45] M. Jacob and P. Landshoff (1980), The Inner Structure of the Proton, *Scientific American* 242 (No. 3, 1980) 46–55.

[46] W. Jentschke, Introduction, *CERN Annual Report 1971* (CERN, Geneva, 1971) pp. 11–26.

[47] W. Jentschke, Introduction, *CERN Annual Report 1972* (CERN, Geneva, 1972) pp. 11–24.

[48] W. Jentschke, Introduction, *CERN Annual Report 1973* (CERN, Geneva, 1973) pp. 11–24.

[49] W. Jentschke, Introduction, *CERN Annual Report 1974* (CERN, Geneva, 1974) pp. 11–24.

[50] B.R. Martin and J. Irvine, Internal Criteria for Scientific Choice: An Evaluation of the Research Performance of Electron High-energy Physics Accelerators, *Minerva* 19 (1981) 408–32.

[51] B.R. Martin and J. Irvine, CERN: Past Performance and Future Prospects I. CERN's Position in World High-energy Physics, *Research Policy* 13 (1984) 186–210.

[52] B.R. Martin and J. Irvine, CERN: Past Performance and Future Prospects III. CERN and the Future of World High-Energy Physics, *Research Policy* 13 (1984) (forthcoming).

[53] W.D. Metz, Fermilab in Transition: The Wilson Era Ends, *Science* 202 (1978) 195–97.

[54] The European Organization for Nuclear Research, *Nature* 184 (1959) 944–45.

[55] New Twist in the Search for Naked Bottom, *New Scientist* 90 (1981) 478.

[56] J. Polkinghorne, The Great Quark Hunt, *The Guardian* (10 July 1980) 13.

[57] A.L. Robinson, New Evidence for Fifth Quark, *Science* 205 (1979) 777.

[58] A.L. Robinson, 1980 Nobel Prize in Physics to Cronin and Fitch, *Science* 210 (1980) 619–21.

[59] A. Salam, Grand Unification of Fundamental Forces, *Science* 210 (1980) 723–32.

[60] H. Schmied, A Study of Economic Utility from CERN Contracts, *IEEE Transactions on Engineering Management* EM-24 (1977) 125–38.

[61] M. Schwartz, Finding the Second Neutrino, in Baggett [3, 41–47].

[62] *Science Citation Index*, produced annually from 1961 to the present by the Institute for Scientific Information, Philadelphia.

[63] S.C.C. Ting, The Discovery of the J Particle: A Personal Recollection, *Science* 196 (1977) 1167–78.

[64] G. Trilling, *Report of the Subpanel on Long-Range Planning for the U.S. High Energy Physics Program of the High Energy Physics Advisory Panel* (Division of High Energy Physics, U.S. Department of Energy, Washington, D.C., DOE/ER-0128, 1982).

[65] L. Van Hove, Review of CERN Scientific Activities, *CERN Annual Report 1976* (CERN, Geneva, 1976) pp. 19–34.

[66] L. Van Hove, The Research Activities of CERN (1976–1980) and the Future of the Laboratory, *CERN Annual Report 1980* (CERN, Geneva, 1980) pp. 27–33.

[67] L. Van Hove and M. Jacob, Highlights of 25 Years of Physics at CERN, *Physics Reports* 62 (1980) 1–86.

[68] R. Walgate, Europe at 400 GeV, *New Scientist* 74 (1972) 217–79.

[69] R. Walgate, Money May Determine West Germany's View, *Nature* 283 (1980) 706.

[70] S. Weinberg, Conceptual Foundations of the Unified Theory of Weak and Electro-Magnetic Interactions, *Science* 210 (1980) 1212–18.

[71] V.F. Weisskopf, Introduction, *CERN Annual Report 1962* (CERN, Geneva, 1962) pp. 16–21.

[72] V.F. Weisskopf, Introduction, *CERN Annual Report 1963* (CERN, Geneva, 1963) pp. 11–15.

[73] V.F. Weisskopf, Introduction, *CERN Annual Report 1964* (CERN, Geneva, 1963) pp. 12–30.

[74] D. Wilkinson, Events Surrounding the Construction of Nimrod, in J. Litt (ed.), *Nimrod: The 7 GeV Proton Synchrotron* (Science Research Council Rutherford Laboratory, Didcot, 1979) pp. 7–20.

CERN: Past performance and future prospects

III. CERN and the future of world high-energy physics

Ben R. MARTIN and John IRVINE *

Science Policy Research Unit, University of Sussex, Brighton BN1 9RF, UK

Final revised version received March 1984

In a series of three papers, we attempt to evaluate the past scientific performance of the three main particle accelerators at the Geneva-based European Organization for Nuclear Research (CERN) over the period since 1960, and to assess the future prospects for CERN and its users during the next ten to fifteen years.

We concerned ourselves in the first paper (Paper I – Martin and Irvine [29]) with the position of the CERN accelerators in world high-energy physics relative to those at other large laboratories working in the field. We dealt primarily with the period from 1969 to 1978, and attempted to establish how the experimental output from the three principal CERN accelerators, *taken as a whole*, compared with that from other major facilities. In undertaking this comparative evaluation, we drew on the method of "converging partial indicators" used in previous studies of three Big Science specialties.

In contrast, the second paper (Paper II – Irvine and Martin [24]) focused in detail on the scientific performance of each of the CERN accelerators *taken individually*. In particular, it asked, first, how the outputs from the CERN 28 GeV (giga or billion electron-volts) Proton Synchrotron compare with those from a very similar 33 GeV American accelerator at Brookhaven National Laboratory over the past two decades. Second, how great have been the experimental achievements of the Intersecting Storage Rings in world terms? And, third, how do the outputs from the CERN 400 GeV Super Proton Synchrotron and from a rival US machine at Fermi National Accelerator Laboratory compare? Attempts were then made to identify the main factors responsible for determining the relative scientific performance of each CERN machine.

These factors are of relevance to the subject of this third paper (Paper III), which sets out to assess the future prospects for CERN and in particular for LEP, the large electron–positron collider scheduled for completion in the latter part of 1988. What are the construction requirements (financial and technical) associated with LEP, and how easily will they be met? How does the scientific potential of LEP compare with that of other major accelerators under construction or planned around the world? In the light of the previous record of the CERN accelerators, to what extent is this scientific potential likely to be realized? What spin-off is there likely to be from LEP to accelerator physics in general? Finally, how "flexible" is LEP – in other words, what is its potential for future development? The paper concludes with a discussion of the extent to which predictive techniques can be utilized in the formulation of scientific priorities, and of the problems in current science policy-making that such techniques might help address.

1. Introduction: The need for predictive science-policy tools

The overriding goal of the various studies of basic science specialties that we have undertaken in recent years [1] has been to develop explicit and systematic methods for evaluating the past scientific performance of major research facilities and their associated user-groups. We have argued that such studies have several important policy applications, particularly in helping to determine when existing experimental facilities have begun to ap-

---

* No order of seniority implied (rotating first authorship). The authors are Fellows of the Science Policy Research Unit, where they work on a range of issues connected with policies for basic and applied research. They gratefully acknowledge the support of the British Social Science Research Council in carrying out this research, and the help so freely given by large numbers of high-energy physicists. Thanks are also due to a number of colleagues at SPRU, especially Professor Keith Pavitt, and to Sir Clifford Butler, Professor John Dowell, and Dr. Owen Lock, for providing useful critical comment on an earlier draft of this paper. However, the conclusions remain the responsibility of the authors alone.

[1] These have been concerned with radio astronomy (Martin and Irvine [28]), optical astronomy (Irvine and Martin [23]), and electron high-energy physics (Martin and Irvine [27]).

Research Policy 13 (1984) 311–342
North-Holland

0048-7333/84/$3.00 © 1984, Elsevier Science Publishers B.V. (North-Holland)

proach the end of their useful research lifetimes. However, in the debate that followed, it became clear that there was also a pressing policy-requirement for improved methods for assessing the likely *future* success of new research facilities.

The need for predictive policy tools is greatest in the capital-intensive "Big Sciences". There are two principal reasons. First, the investments involved are now of such a magnitude that discussions of them are central in the formulation of overall national science policies. The cost of the planned new LEP accelerator at CERN (over 900 million Swiss francs) is, for example, appreciably greater than the total British Science and Engineering Research Council budget for basic natural science in 1983/4. At a time when there are severe pressures on funding, it is important that investment decisions are taken on the basis of the fullest possible information, particularly in Big Science where such decisions clearly have major implications for the funding of other specialties.[2]

Second, there are reasons for questioning the extent to which the scientific communities in Big Science specialties are still able to make investment decisions solely, or even largely, on the basis of perceived scientific merit (cf. Irvine and Martin [26]). As scientific activity in such specialties is concentrated in ever fewer research centres (for example, in the United States there are in 1984 only four experimental high-energy physics centres, and in Western Europe just two), so decision-making is increasingly influenced by institutional and political pressures. One possible consequence is that new facilities may not be sited at the research centres best able to exploit them to the full. The underlying problem is that it is becoming more difficult in Big Science to locate neutral peers capable of providing sufficiently disinterested judgements; all potential peers tend either to have some professional interest in a proposed new project, or to be associated with a competing set of interests which would benefit from a negative decision on that project.

The emergence of such imperfections in the peer-review process (the method traditionally used in allocating resources for basic science) has increased the need for other sources of information to aid policy-makers in determining future priorities – particularly where the decisions involve the distribution of funds between specialties as they do indirectly in the case of Big Sciences. In this respect, we would argue that systematic comparisons of past scientific performance (of the sort reported in Papers I [29] and II [24]) can be an important input into decision-making, particularly when generated by analysts outside the social structure and reward system of the reseach community concerned. When resources are heavily concentrated on a single centre or research facility, it becomes all the more important to ensure that they are used effectively and that there is some monitoring procedure for quickly identifying problems limiting the scientific output. Equally important, when centres apply for funds to replace obsolete equipment (for example, an accelerator or telescope), it is desirable to know how successfully the previous research facilities have been operated, in particular compared to similar facilities at rival centres. "External" evaluations are probably more likely to be trusted by sections of the scientific community outside the Big Science concerned; while they may accept the need for major investment in capital-intensive research facilities, they clearly require some assurance that such facilities are actually producing important scientific results or are likely to do so in the future.

This said, it is clear that "track record" is only one element, albeit an important one, in helping to predict likely future research performance. Although perhaps marginally better than no forecasts at all, predictions of the future based upon simple extrapolations from the past are unlikely to be particularly successful (see Miles and Irvine [32]), especially in areas of science characterized by rapid change. This is not to say that extrapolations should never be attempted, but rather that they are more likely to have some validity when based upon an understanding of the factors that have structured past performance, and of the extent to which they are likely to affect future performance. For this reason, an attempt has been made in the present study to extend the previous evaluation methodology (which was concerned with assessing the past performance of various Big Science facilities – see footnote 1) to include an examination of the reasons why the research facilities in question (high-energy physics accelerators) have performed with greater or lesser success (see in particular tables 9, 11, and 12 in Paper II [24]). A range of information has thus been obtained concerning the factors structuring success and failure in the past – for example, whether there has been a strong user-group associated with the accelerator, whether the scientific management of the facility has been effective, or whether from a technical viewpoint the accelerator and subsidiary instrumentation have been of a sufficiently high quality.

---

[2] For example, Britain spent £83.2 million in supporting basic research in astronomy, space, and nuclear physics in 1981/82. This compared with £21.8 million spent by the Science and Engineering Research Council on peer-reviewed grants for all other areas of natural science, £20.7 million by the Social Science Research Council, £42.1 million by the Agricultural Research Council, £54.3 million by the Natural Environment Research Council, and £101.7 million by the Medical Research Council (see Irvine and Martin [26]).

Yet even assuming that these factors continue to be important, this is still an insufficient basis on which to make decisions concerning a major new facility. What is needed in addition is information relating to the characteristics of the proposed new facility and the research to be carried out on it – the "ripeness" of the research area, the relation of the instrument to other facilities (for instance, whether it duplicates, or is complementary to, facilities elsewhere), the degree of technological "risk" associated with the instrument (for example, whether it relies on a new and untested technique), and the likelihood of "spin-off" in the form perhaps of instrumental techniques that can be applied in other research activities.

In this way, then, it is possible to envisage the formulation of a set of criteria to help assess the merit of proposals for future research facilities. Certainly, many of these criteria are already used informally by the scientific community. However, such criteria are not always made publicly explicit, and the information required to utilize them tends to be accessible *only* to researchers within the specialty concerned. Hence, in specialties where there is a high degree of concentration of research efforts (and a resulting formation of strong interest-groups focused on each of the main centres in the field), there is a danger that the peer-review process may be reduced to little more than a battle between institutional interests (cf. Irvine and Martin [26]) The underlying problem is that, in the present system of decision-making, there are inadequate mechanisms for ensuring accountability to those *outside* the Big Science concerned, even to scientists in neighbouring specialties. The aim in this paper is to consider the possibilities for developing a framework to be used in analyzing future policy options in a systematic and publicly accessible way. If the scope for institutional lobbying is to be limited, then deliberations over costly new research instruments like accelerators or telescopes must be open to scrutiny by a wider body of scientific and public opinion. This would *not* remove ultimate responsibility for the distribution of resources for research from scientists, but merely ensure that scientific decision-making was conducted in a more public arena.

In this paper, we attempt to assess the future scientific prospects for CERN. In particular, we consider whether the set of criteria described below can aid analysis of the likely prospects for CERN's major new accelerator, LEP, relative to those of accelerators elsewhere. If so, to what extent do the results throw into question the adequacy of existing decision-making procedures for determining the future priorities of a major laboratory like CERN? This is especially pertinent since LEP marks something of a departure from previous construction projects in terms of the long-term financial commitment it implies for CERN Member States and its delimitation of future options for the laboratory. Whereas previous accelerator construction programmes at CERN have been essentially discrete efforts with a time-horizon of perhaps six years from inception to commissioning, the LEP project is rather different. Besides Phase I of the project (to reach a beam-energy of 50 GeV) scheduled for completion in the second half of 1988, subsequent phases to reach first approximately 90 GeV and then 130 GeV are also planned, each of which may take several years to complete. Formal approval of these later phases will obviously not be obtained for some time, but there must be confidence at CERN that, once Phase I and the 27-kilometre tunnel it requires have been completed, the arguments for continuing to increase the energy up to 130 GeV will be exceedingly strong. As with Phase I, high-energy physicists will probably be able to argue that these later phases can be carried out without any overall increase (in real terms) in the CERN budget. At such a rate of funding, a 130 GeV version of LEP would probably not come into operation until the early 1990s, i.e. nearly ten years from now.

At that stage, there are likely to be strong arguments that, since the 27-kilometre tunnel already exists, it would be relatively cheap to construct a superconducting proton synchrotron of perhaps up to 10,000 GeV (10 TeV) by placing it alongside LEP. Such a machine could take another five years or so to construct. Moreover, once completed, it might be converted into a large proton–antiproton collider; and, in conjunction with LEP, it could also be used to generate high-energy electron–proton collisions. These latter two projects might each take a further three or four years to bring to fruition. In this way, LEP will provide CERN with a succession of possible new projects well into the first decade of the next century.

Since the choice to embark on LEP will have implications for what CERN is doing twenty years or more from now, it is surprising that there was not more *public* discussion of LEP, and especially of its significance for the future development of the laboratory, before the formal go-ahead for the project was given in 1981. Certainly, extensive discussions took place *within* the high-energy physics community. In addition, there were negotiations between high-energy physicists and research-funding agencies in the Member States. However, these discussions were circumscribed by the decision of CERN management that LEP should be treated not as a "new" accelerator but as an extension of existing CERN facilities, a decision taken to minimize the risk that political debate might delay the project's start. LEP may indeed turn out to be the best accelerator for CERN, but what is not clear is whether the fairly heavy reliance on the internal debates of a scientific community whose research interests centre largely around a single laboratory forms an adequate basis for committing substantial long-term resources – particularly when the funding involved is of such a magnitude that it cannot fail to have implications for other scientific specialties.

In analyzing whether improved procedures can be developed for assessing the future prospects of large capital projects, we shall first examine the history of the LEP project. This is followed by a brief review of the other main accelerator (or collider [3]) projects planned or underway around the world. The central sections of the paper then draw certain comparisons between these projects. The first focuses on their relative financial and technical requirements, and attempts to assess the extent to which these are likely to affect their completion within the currently planned schedule. The second examines the scientific potential of each new accelerator, considering in turn whether it will have a world lead in a new energy region (and, if so, for how long), the likely capacity of that accelerator to generate new physics results, and the ability of its user-community to exploit the scientific potential of the new facility to the full. The third and final set of comparisons centres on the spin-off, or contribution to accelerator physics and technology, likely to be generated by each machine, and on its potential for development at a later stage to yield a new research facility. The results of the comparisons along these three dimensions are then used to arrive at an overall assessment of the future prospects for CERN over the next ten to fifteen years. The paper concludes with a brief discussion as to the likely utility of the methodological approach described below for science-policy purposes, particularly in those fields of science characterized by heavy capital expenditure on centralized research facilities.

Given the relative lack of experience that exists with predictive science-policy analysis, this attempt at constructing a framework for systematically assessing future prospects should be regarded as no more than a provisional first step, designed to demonstrate – in particular to senior policy-makers concerned with the overall distribution of research funds, but also to the scientific community at large – the potential utility of such external evaluations. The intention is that such assessment excercises be undertaken prior to major research investment decisions, and then repeated periodically during the construction and operation of a facility. It should also be stressed that studies such as this which are concerned with the future may "date" quite rapidly as they are overtaken by unforseen events. Nevertheless, while questions of detail may change appreciably in the period between the writing of this paper (in late 1983) and its publication and circulation, we believe most of the main issues and trends will not. To this extent, there is therefore an element of "testability" about the validity of the approach that we have adopted.

---

[3] In this and Papers I [29] and II [24], the term "accelerator" is sometimes used generically to cover both accelerators and colliders.

2. The historical background to the LEP project

Until the start of the 1970s, the history of experimental high-energy physics had been largely dominated by the contributions of proton accelerators. With the exception of deep-inelastic scattering, which was discovered on the Stanford (electron) linear accelerator in 1968 (see Paper II [24]), virtually all the major advances in the field over the previous twenty years had come from proton accelerators – generally from the highest-energy machine operating at the time. Even in 1970, most senior high-energy physicists saw no obvious reason why this trend should not continue, and the thoughts of many of the more ambitious accelerator-builders began to turn to proton synchrotrons with an energy of 1000 GeV (1 TeV), and to proton–proton colliders with a beam-energy an order of magnitude higher than the 30 GeV of the CERN ISR, then nearing completion. Over the next few years, examination of these possibilities advanced furthest at Brookhaven in the United States and at Serpukhov in the Soviet Union, both of which were considering replacements for their existing facilities. Less progress was made at the two other major world proton centres, CERN and Fermilab (near Chicago), where scientists were preoccupied with bringing into operation and exploiting their new accelerators (the ISR, SPS, and 400 GeV machines).

Then, in 1974, high-energy physics was shaken by the so-called "November Revolution."[4] The discovery of the J/psi (followed quickly by those of other related particles) not only completely reoriented the direction of mainstream thinking in particle physics; it also had a dramatic impact on the plans of accelerator-designers. Suddenly, the previously "unfashionable" technology of the electron–positron collider was thrust into the limelight as experimentalists in the United States using the Stanford accelerator, SPEAR, unearthed a succession of crucial discoveries – the psi-prime, the heavy lepton tau, charmed mesons, and so on – that until then had eluded the much higher-energy proton machines. One fairly immediate consequence was the race between the German laboratory, DESY (the largest of the European electron accelerator centres) and Stanford to construct a higher-energy electron–positron collider,[5] with the latter quickly dropping its original plans for a proton–electron facility. A longer-term outcome was that West European physicists began in the mid-1970s to consider the possibility of constructing an enormous 100 GeV electron–positron storage-ring, a proposal which eventually materialized as the LEP project.

Besides the discovery of the J/psi, there was another notable advance within high-energy physics at about the same time that came to exercise a significant influence on the thoughts of accelerator physicists. This was the considerable success of the Weinberg-Salam unified theory of electromagnetic and weak interactions, in particular the confirmation at CERN in 1973 of the theory's prediction of neutral currents (see Paper II [24]). However, the findings of this and subsequent experiments were regarded by physicists as giving only corroborative rather than conclusive evidence in favour of the theory; the crucial test was whether the "intermediate vector bosons" (the neutral Z_0, and the charged W^+ and W^- particles) postulated by the theory as the carriers of the weak force could actually be detected. The eventual formulation of a theory unifying the four types of force encountered in nature has been aptly characterized as "like the Holy Grail to particle physicists" (Robinson [34, p. 192]), and the discovery of the predicted Z and W particles rapidly came to be seen as the vital first step towards that elusive goal. The problem, however, was that the predicted masses of these two types of particle (approximately 90 and 80 GeV respectively) required an accelerator producing a centre-of-mass energy of nearly 100 GeV. This was an energy region far beyond the reach of accelerators then operating in the mid-1970s. As it became clear that it might take up to ten years to construct a purpose-built machine (like ISABELLE in the United States or LEP) capable of attaining this energy, certain physicists began to consider in 1976 whether existing accelerator facilities might be modified to bring the new particles within range somewhat sooner. From this grew the idea for converting the CERN SPS into a proton–antiproton collider. Since the history of this project is closely interwoven with that of LEP, it is worth looking at it briefly.

---

[4] This and the other dramatic discoveries made in the subsequent three years are described in detail in Paper II [24].

[5] Though the Stanford electron–positron collider is named PEP, the first "p" in the acronym originally stood for "proton". See Metz [30, p.853] for details of this earlier project.

During the mid-1970s, concern was being expressed within CERN that the SPS was coming into operation over four years behind the virtually identical Fermilab accelerator (cf. Van Hove [41, p.31]). Hence, when the Italian physicist, C. Rubbia, began to advocate the conversion of the SPS (and the Fermilab accelerator) into a proton–antiproton collider capable of reaching the intermediate vector boson region, he found a receptive audience within the European high-energy physics community. A further incentive to proceed with the project came with the realization that the SPS had been so constructed that it could be converted relatively cheaply and – perhaps of greater importance to CERN – far more quickly than the Fermilab accelerator, giving European physicists the chance of turning the tables and gaining several years lead over their American rivals. One further factor that may have contributed to the attractiveness of this project concerned its role in the longer-term budgetary plans of CERN. In the absence of a major capital-construction project to fill the hiatus between the completion of the SPS in 1976 and commencing the building of LEP in the early 1980s, CERN would have found it politically difficult, if not impossible, to prevent its annual budget being cut from the "ceiling" figure of about 600 million Swiss francs (MSF) to something closer to the "base load" value of approximately 400 MSF required for experimental operation and routine investment purposes. Once reduced to the "base load" figure, it might have proved even harder for CERN to convince all the Member States at a later date that the budget needed to be substantially increased again in order to begin the construction of LEP.[6] The proton–antiproton collider, with its relatively modest cost (200 MSF) for a new high-energy physics facility, neatly filled the five-year "gap" between the completion of the SPS and commencement of investment on LEP.[7]

The main uncertainty associated with the proton–antiproton collider concerned the technical difficulty of achieving sufficient luminosity (one of the factors determining the number of particle collisions per second) to make the physics accessible. However, the results of various tests of the technique of stochastic cooling (pioneered at CERN) between 1976 and 1978 were sufficiently encouraging to suggest that a luminosity of 10^{30} cm^{-2} s^{-1} (the eventual design figure) might be achievable – sufficient to generate tens of Z_0 particles and hundreds of W particles per day. Even if a luminosity of approximately an order of magnitude less than this were achieved, it was assumed that this would still be high enough to permit the new particles to be discovered some years ahead of ISABELLE, the Brookhaven proton–proton collider then scheduled to begin operation in about 1983. Even so, the proposed collider was not without its opponents; some users of the SPS accelerator in particular were alarmed that the conversion of the machine into a collider would cause the loss of up to a year from the fixed-target experimental programme (which was already labouring under the handicap of having started over four years behind that of Fermilab), that it might affect the reliability of the original synchrotron (as it in fact did), and that, when completed, the collider would occupy a significant proportion of the SPS's time, leaving insufficient for fixed-target experiments. Nevertheless, the CERN management decided in 1978 that this major sacrifice of prime research time was a price worth paying in return for the chance of discovering the all-important intermediate vector bosons. Once the go-ahead had been given, construction of the collider proceeded rapidly, the first collisions being observed in July 1981, and, as we saw in Paper II [24], the W and Z intermediate vector bosons were discovered amid much publicity in 1983.

In the meantime, there had been significant progress with the LEP project. Between 1975 and 1978, considerable debate had taken place within the West European high-energy physics community over the facilities that would best meet their experimental needs in the period up to the end of this century. Among the options considered for CERN were a very high energy fixed-target proton synchrotron, a proton–proton collider, an electron–proton collider, and a large electron–positron collider, LEP. The European Committee for Future Accelerators (ECFA) played a prominent role in this discussion (much of it was carried out in the closed sessions of the "Restricted ECFA"), and eventually in 1977 came down strongly in favour of LEP. Their decision appears to have been based on a widely accepted set of scientific arguments – in particular the successes of existing electron–positron colliders in contributing to the "new physics" of quarks and leptons, and the prediction by a large body of theorists that the energy of the Z_0 was such that LEP would be the ideal research instrument on which to mount a systematic investigation of the particle's properties (its width, decay-channels, and so on). Another strongly influential factor was the notion of complementarity: Fermilab and Serpukhov were planning proton synchrotrons of 1 TeV and 3 TeV respectively; and Brookhaven was about to embark upon the construction of a large proton–proton collider. By opting for LEP, so the argument went, CERN would be guaranteed a unique experimental facility, thus avoiding the direct competition that had limited the relative scientific impact of both the PS and SPS (see Paper II [24]). A final factor that appears to have influenced the choice of LEP was the potential of such a facility for future development – the so-called "real-estate argument" – a feature of the decision we shall discuss later.

---

[6] Certainly, CERN would then have found it difficult to argue that LEP constituted no more than an "extension" to the existing experimental programme (a point discussed further below).

[7] Similarly, one of the main arguments for proceeding with the construction of the ISR in the late 1960s (according to senior physicists interviewed at CERN) had been to fill the "gap" between refurbishing the PS and beginning work on the politically delayed, but higher priority, SPS.

By 1978, the *CERN Courier* [7, p.431] was able to report that, after much discussion within the European high-energy physics community,

> The consensus has come out strongly for an electron–positron machine to take colliding beam energies well beyond those which are accessible with PETRA at DESY and PEP at Stanford.

The main technical parameters of the machine had also been determined by this time (in particular, a ring of 30 kilometres in circumference had been agreed, rather than the 50-kilometre and 20-kilometre options considered previously), while cost estimates were calculated shortly afterwards. By adopting a "missing-cavity design" (in which radio-frequency cavities were progressively added), it was proposed to build LEP in several stages, reaching a beam-energy of 50 GeV for a cost of just over 900 MSF. Energies of 65 GeV and 90 GeV would subsequently be achieved for a total cost of 1065 MSF and 1275 MSF respectively (cf. [8, pp.10–11]). The final design figure of approximately 130 GeV would be reached by replacing the conventional radio-frequency cavities with elements based on new superconducting technology.[8]

Agreeing upon the design and costings for a new project is one thing; obtaining the funds and political agreement from Member States to go ahead with the construction programme is another, as we saw in the discussion of the history of the SPS project (see Paper I [29]). The SPS was severely delayed because of lack of agreement over both finances (with Britain threatening withdrawal) and the site of the new accelerator. According to senior physicists whom we interviewed, similar political infighting among Member States began to develop over the LEP project, with strong pressure being exerted in particular by the Germans to site the new accelerator in the Federal Republic. It also soon became apparent from preliminary discussions with funding bodies in the various Member States that it would be politically difficult, if not impossible, to finance the new accelerator by an increase in the overall CERN budget. With the experience of the SPS behind them, the CERN management this time took clear and decisive action. First, they decided to propose to Member States that the new machine be financed from the existing CERN budget (some 600 MSF per annum at 1978 prices), even though this would entail closing the ISR[9] and perhaps the small synchrocyclotron, and curtailing the SPS experimental programme. This also involved trimming the costs of LEP, in particular by using the SPS to inject electrons and positrons into the main ring rather than building a new 22 GeV electron synchrotron. Although this further reduced the SPS programme by up to 10 percent, it slightly decreased the financial and manpower commitments for LEP. Second, it was decided that LEP could now be treated as "an extension of existing CERN facilities" [9, p.192], to be absorbed under the "basic programme" of CERN, rather than as a new accelerator [10] requiring the authorization by Member States of a supplementary programme – authorization which might not be immediately forthcoming. Third, in response to worries by the smaller countries about possible cost over-runs in the LEP project as a result of technical problems, [11] CERN had to agree that, in future, any proposed real increase in the laboratory's annual budget could be vetoed by a single Member State, instead of the previous situation where agreement could be obtained by a simple two-thirds majority (cf. Walgate [42, p.275]). This means that any unforseen difficulties requiring a major increase in expenditure will have to be met by either postponing the date of the first experiments on LEP or curtailing even further the existing research programmes on other facilities. We shall return to discuss the significance of this later. Although there were initially certain reservations among three of the Member States about the effect of LEP on other experimental programmes, by the end of 1981 all twelve had agreed to proceed with the project. As matters stand at the time of writing, it is anticipated that Phase I (to achieve a beam-energy of 50 GeV and to construct four out of the eventual eight experimental halls) will be completed by the second half of 1988 (cf. [22, p.228]). The estimated cost in 1981 prices is 910 MSF; to this must be added between 250 and 280 MSF for building the four planned detectors (cf. Robinson [38, p.722]), of which CERN will contribute about 60 MSF, the remainder being provided by the national funding agencies which will support the research activities of the predominantly university-based LEP users (including many from non-Member States of CERN).

---

[8] Robinson [35, p.531] has quoted an approximate figure of 400 MSF for the cost of replacing the conventional cavities by superconducting elements and for construction of the four remaining experimental areas. This project is referred to below as LEP-130.

[9] This decision resulted in a vocal campaign by a significant minority of CERN physicists against closure of the ISR. Arguing that the machine had a good track record and was still a unique facility, many physicists claimed in interviews with us that, when their support for LEP was first sought, it was not made clear that proceeding with this project might involve the closure of the ISR. Instead, many would have preferred to see the LEP project spread over a longer period, thus freeing resources for continuing the ISR programme. Some admitted that closure of the ISR was the political sacrifice that probably had to be made to guarantee the future for LEP, but were bitter about the lack of public discussion on the matter. In a similar vein, the Scandinavian countries rapidly mobilized against the possi-

3. Likely competitors to LEP

Before we move on in subsequent sections to assess the comparative advantages and disadvantages of LEP, it is first necessary to examine the main characteristics of potential rival accelerators currently in operation at other laboratories or planned for construction over the next decade. These competitor machines have been classified into four categories: (1) fixed-target proton synchrotrons; (2) proton colliders; (3) electron–proton colliders; and (4) electron–positron colliders.

3.1. Fixed-target proton synchrotrons

Details are given in table 1 of the three planned (or possible) major new accelerators which may be commissioned before the end of the century. For comparative purposes, similar data are also given for the largest proton synchrotrons already operating.

3.1.1. 3 TeV UNK (Serpukhov, USSR)

As was seen in Paper II [24], by the time the experimental programme of the Serpukhov 70 GeV accelerator was in full operation, the much

---

ble closure of the synchro-cyclotron (which many of their physicists use) and eventually obtained agreement for the continuation of the experimental programme on this accelerator on a limited basis.

[10] This would seem to be a somewhat artificial distinction and one that might prove rather hard to defend if subjected to wider public scrutiny. According to this definition, there have been virtually no major "new accelerators" since 1972, nor are any planned: the ISR and SPS were merely "extensions" of the CERN PS, SPEAR and PEP were extensions to the Stanford linear accelerator, ISABELLE would have been just an extension of the Brookhaven AGS, and so on. While this reclassification of the status of LEP may have been admirably suited to CERN's purposes in that it greatly reduced the risk of political disagreement among Member States over the project, it constrained the options of Member States as regards discussions of other types of new facility. It might have been expected that such a terminological "sleight of hand" would give rise to some controversy about high-energy physicists' arrogation of government funds while seeking to limit the degree of public accountability, but none seems to have arisen.

[11] Problems were, for example, anticipated with tunnelling under difficult rock conditions in the Jura mountains. As a result of testborings, the precise location of the LEP tunnel had to be changed: the circumference of the accelerator was decreased from 30 to 27 kilometres and the length of the tunnel under the mountains reduced from 12 kilometres to first 9 and then 3 kilometres (see Walgate [42, p.275]). Severe construction problems still potentially remained to be overcome, however.

Table 1
Major proton synchrotrons – present and proposed [a]

	Date of first experiments	Maximum beam-energy (GeV)	Centre-of-mass energy (GeV)	Intensity (particles per pulse)	Estimated cost (MSF [b])	Period of world leadership (years)	Future options [c]
Serpukhov	1967	76	12	5×10^{12}	–	5	–
Fermilab	1972	500 [d]	31	2×10^{13}	–	4	–
CERN SPS	1976	500 [d]	31	2×10^{13}	–	8	–
Fermilab 1 TeV	1983	1000 [e]	43	3×10^{13}	~ 250?	~ 6?	p$\bar{p}$, pp
Serpukhov UNK	1990?	3000	75	~ 10^{14}?	~ 400?	~ 5?	pp, p$\bar{p}$
CERN Superconducting Synchrotron?	mid-1990s?	10000?	135	?	?	?	p$\bar{p}$, pp, ep

[a] This lists only the world's highest-energy machines.

[b] These are the approximate estimated costs in millions of Swiss francs based on 1983 exchange rates. (An exchange rate of $1 = 1 Soviet rouble has been used.)

[c] pp = proton–proton collider; p$\bar{p}$ = proton–antiproton collider; ep = electron–proton collider.

[d] This figure is rarely achieved. The "normal" maximum energy is 400–450 GeV.

[e] Experiments at an energy of about 800 GeV are planned for 1984.

higher-energy machine at Fermilab was nearing completion. Once their world lead in energy disappeared in 1972, Soviet high-energy physicists began planning for the construction of a major new proton synchrotron, UNK, the energy of which was originally set at 2000 GeV (2 TeV). It was reported in 1976 that the Soviet Government had pledged 200 million roubles (about $200 million) to the project [4, p.401]. By 1980, the final design parameters had been settled and the proposed energy raised to 3 TeV.[12] It was originally estimated that the machine would take approximately seven years to complete (Wilson [43, p.40]) from the time construction began in 1981, although 1990 is now considered to be a more realistic completion date (cf. [45, p.301]).

3.1.2. 1 TeV Doubler / Saver (Fermilab, USA)

Shortly after completing its 400–500 GeV accelerator in 1972, Fermilab began work on the Energy Doubler project to increase its energy to 1000 GeV (1 TeV) by installing superconducting magnets. As with LEP, the history of decision-making over this new accelerator has been rather complex. The project was initially classified by Fermilab as an internal R&D programme, for which no special authorization was required; at this stage, the cost was roughly estimated at $35 million [3, p.13], which the laboratory management felt could be found from within its general operating budget. In 1977, however, the United States Office of Management and Budget insisted that the work had to be defined as a construction project. Since the procedure for obtaining authorization for projects costing under $50 million is considerably less severe than for those requiring higher expenditures, efforts were made to keep the estimated costs below this limit. When it eventually became clear that this was no longer feasible (in part because of greater than expected technical problems with superconducting magnets), the project was split into two parts: (1) the Energy Doubler (or Energy Saver [13]) – the construction of a "bare" superconducting magnet ring; and (2) the Tevatron – the provision of all the subsidiary instrumentation to make possible fixed-target experiments at an energy of 1 TeV, and proton–antiproton collisions at a centre-of-mass energy of 2 TeV. Furthermore, when the estimated costs of the Tevatron in turn rose above the $50 million mark, this project too was divided into two parts – Tevatron I and Tevatron II. By 1981, the estimated costs of constructing the 1 TeV fixed-target accelerator (Tevatron II) had risen to between $110 and $120 million,[14] to which must be added the cost of upgrading the beams and experimental areas for the higher-energy experiments. Experiments at an energy of between 800 and 900 GeV are to be run in 1984, with the maximum energy being achieved a year later (cf. Robinson [39, p.816]). It will probably then remain the world's highest-energy accelerator for the rest of the 1980s.

---

[12] The existing 76 GeV accelerator, suitably upgraded, will be used to inject protons into a conventional 400 GeV booster synchrotron, and subsequently into a 3 TeV superconducting-magnet proton synchrotron (cf. [11, p.147]).

[13] Although Fermilab's primary consideration was to build a 1 TeV accelerator, the Department of Energy (the main funding body in the US for high-energy physics) was more concerned that the laboratory should make efforts to reduce its electricity bill, a goal towards which superconducting magnets would contribute: hence the alternative project title, the Energy Saver.

3.1.5. 10 TeV Superconducting Synchroton (CERN)

The 27-kilometre tunnel for LEP is designed to be large enough to accommodate one or more additional rings of magnets. Consideration has already been given at CERN to the possibility of installing a ring of high-field superconducting magnets (fields of up to 10 tesla should be feasible by the mid-1990s) to yield a proton synchrotron with a maximum energy of approximately 10 TeV (cf. Wojcicki [44, p.26]).

3.2. Proton colliders

Details are given in table 2 of the main proton (and antiproton) colliders likely to become operational over the next decade; as before, the table includes comparative data on currently existing facilities. In addition, information on the proposed Colliding Beam Accelerator (formerly called ISABELLE) is included, even though this project was halted in 1983, the reason being to ascertain the likely prospects for this collider had it been completed.

3.2.1. 400 GeV Colliding Beam Accelerator / ISABELLE (Brookhaven, USA)

When Brookhaven completed the AGS Conversion Programme (see Paper II [24]) in 1971, the laboratory began to plan a superconducting proton–proton collider with a beam-energy of 200 GeV. Early tests with the prototype superconducting magnets proved very encouraging;[15] the field-strength of 4 tesla required for a 200 GeV machine was easily attained, and, by 1976, a figure of 5 tesla seemed to be within reach. The US High Energy Physics Advisory Panel (HEPAP) was pleased with the progress achieved in this technologically demanding development programme, and, urged on by theorists eager to see experimental data from even higher energies, recommended that the planned beam-energy be increased to 400 GeV. With hindsight, this decision is now regarded by many physicists as a mistake. Although an industrial contract was awarded in 1978 for winding the magnet coils, by 1980 it was clear that the magnet design (the so-called "braid" magnet) chosen by Brookhaven was incapable of achieving the necessary field-strength. In 1981, after considerable internal organizational upheaval at the laboratory, these plans were finally dropped to be replaced by a design much closer to that being developed at other accelerator centres, including Fermilab.

Inevitably, these severe technical problems resulted not only in a dramatic escalation in the estimated costs of ISABELLE, as it was then called, but also in the likely completion date being pushed ever further into the future. In 1976, the cost of the 200 GeV collider was estimated as $166 million, and a year later, after the energy had been increased to 400 GeV, a figure of $245 million was quoted, together with a likely completion date of 1983. Construction of the accelerator tunnel started later that year, and, by the time the original magnet design was scrapped, a total of $130 million had already been spent (cf. Broad [1, p.1089]). By 1982, it was clear that the collider would not be completed until 1988 and that a further $500 million would be required (cf. Trilling [40, p.26 and p.49]). In view of the parlous state of US high-energy physics funding, serious doubts were expressed about whether, despite the major expenditure already incurred, it was worth completing ISABELLE (cf. [40, p.50]).

In response to this situation, various alternatives to ISABELLE were considered, including an electron–proton collider and a heavy-ion collider. However, the cost-savings associated with these options were found to be modest and were deemed insufficient to justify the significant reduction in physics potential associated with them. As a result, the Brookhaven management decided in February 1983 to persevere with the original plans for a

---

[14] Figures provided to the authors during a site-visit to Fermilab in late 1981.

[15] As early as 1973, it was reported that, "The performance of the [prototypes] seems to have answered several outstanding questions concerning superconducting magnets. One concerns the reproducibility of magnets." [2, p.374]. In the light of the major problems with magnets experienced later (described below), this conclusion seems to have been a little premature.

Table 2
Major proton colliders – present and proposed

	Date of first experiments	Maximum beam-energy (GeV)	Centre of mass energy (GeV)	Luminosity ($cm^{-2}s^{-1}$)	Estimated cost (MSF [a])	Number of experimental areas	Period of world leadership (years)	Future options
CERN ISR	1971	31	62	10^{32}	–	8	10	–
CERN p$\bar{p}$ [b]	1981	270	540	10^{30}? [c]	200 [d]	2	5	–
Fermilab p$\bar{p}$ [b] Collider	1986	1000	2000	10^{30}?	~ 350? [d]	2	~ 5–10?	–
Brookhaven ISABELLE [e]	1987/8	400	800	~ 10^{33}?	~ 1700?	6	0	ep, heavy-ion collider
Serpukhov UNK	early 1990s?	400 + 3000	2200	~ 10^{32}?	?	?	?	3000 + 3000 GeV $\bar{p}$p and pp
Fermilab Dedicated Collider?	early 1990s?	2000	4000	~ 10^{31}?	~ 900?	6	?	ep
CERN Superconducting Collider?	mid-1990s?	10,000	20,000	?	?	?	?	ep
US Superconducting Super Collider?	mid-1990s?	20,000	40,000	~ 10^{33}?	6000–8000? [d]	?	?	?

---

[a] See note b to table 1.

[b] Proton–antiproton collider.

[c] This is the design-figure. The highest value obtained up to the end of 1983 was 2×10^{29}.

[d] This includes the costs of the detectors.

[e] Project discontinued in 1983.

253.

proton–proton collider (now renamed the Colliding Beam Accelerator or CBA) (cf. [21, p.127]). In an effort to regain the support of the US high-energy community for the project, an accelerated construction schedule enabling the first colliding beams to be obtained in October 1987 was drawn up (cf. Wojcicki [44, p.31]). However, in July 1983, HEPAP recommended that the CBA be discontinued (cf. [44, p.7]) and the project was finally halted by the Department of Energy a few months later. By then, a total of some $200 million had been spent on the accelerator.

3.2.2. 1 TeV proton–antiproton collider – Tevatron (Fermilab, USA)

The Fermilab proton–antiproton collider project began in late 1976, replacing earlier plans for a proton–proton machine. As with the laboratory's 1 TeV fixed-target accelerator, the projected cost has risen appreciably over time – from an initial value of some $40 million to an estimated $70 million in 1981. In addition, the decision in 1982 to develop a new and more efficient antiproton source was reported as increasing the cost by a further $40 million. To this must be added the outlay on the two detectors, giving an overall total of around $170 million.[16] The first proton–antiproton experiments are planned for 1986, which will give CERN five years lead with its (lower energy) collider – the same advantage previously enjoyed by Fermilab with respect to the 400 GeV proton synchrotrons.

3.2.3. Serpukhov collider (USSR)

Once the 3 TeV proton synchrotron (UNK) has been completed at Serpukhov, it is intended to collide 3 TeV protons with 400 GeV protons from the booster synchrotron (see footnote 12 above), giving a centre-of-mass energy of approximately 2200 GeV. There are also preliminary plans for phase II of UNK which include 3 + 3 TeV proton–antiproton colliding beams and, with the addition of a second superconducting-magnet ring, 3 + 3 TeV proton–proton collisions.

3.2.4. 2 TeV Dedicated Proton–Antiproton Collider (Fermilab, USA)

In 1983, Fermilab put forward a proposal to build a proton–antiproton colliding facility (called the Dedicated Collider) with planned beam-energies of up to 2 TeV and a design luminosity of over 10^{31} cm^{-2} s^{-1}. Although it would use the 1 TeV Tevatron as an injector, it would have the advantage of freeing the Tevatron to run almost entirely for fixed-target physics. The necessary preparatory R&D and construction were estimated as likely to take some six years to complete, and to cost between $380 million and $450 million (cf. Wojcicki [44, p.34]). However, the US High Energy Physics Advisory Panel (HEPAP) decided in July 1983 that Fermilab should not proceed with the Dedicated Collider for the time being, and that efforts should instead be concentrated on the Superconducting Super Collider (see below).

3.2.5. CERN Superconducting Collider

As already mentioned, the tunnel for LEP can accommodate one or more additional rings of magnets. By installing a ring of high-field superconducting magnets, a proton–antiproton colliding-beam facility with beam-energies of up to 10 TeV could be constructed at CERN. It has been estimated that, with appropriate funding, such a facility could be built over a period of about five years (cf. [44, p.26]).

3.2.6. Superconducting Super Collider (SSC) (USA)

The year 1983 marked something of a turning point in US particle physics. After strong pressure from the Administration, the High Energy Physics Advisory Panel (HEPAP) was finally able to reach a consensus agreement that the Colliding Beam Accelerator project at Brookhaven be discontinued. According to those interviewed by us, this agreement would never have been reached had high-level promises not been forthcoming that support would be likely for an even larger project. As a result, frantic activity took place in early 1983 to discuss possibilities for the new facility. A decision was soon made to opt for an accelerator/storage-ring complex to accelerate protons to an energy of 10 to 20 TeV, and to collide them. Outline plans for the facility were discussed at a workshop held at Cornell University, HEPAP unanimously endorsing in July 1983 (three months later) a recommendation by its Sub-panel on New Facilities for the immediate initiation of the project. Provisional estimates suggest that the construction of such a machine together with the necessary preliminary

---

[16] Robinson [39, p.816] quoted a figure of $300 million in 1982 for the combined cost of the 1 TeV accelerator (including its upgraded experimental areas) and the collider

R&D could take between nine and fifteen years and cost a total of up to $3 billion (cf. [47, p.390]), while four or five detectors could cost as much as $ 1 billion (cf. [46, p.19]). Budgetary plans have already been laid for a preliminary R&D phase for the project. If the United States then decides to proceed further, the Department of Energy annual budget for high-energy physics would have to be approximately double the 1983 figure during the five to six years of construction.

3.3. Electron–proton colliders

Nearly every major accelerator centre has at one time or another considered building an electron–proton collider, although this has tended to be regarded as a possible second stage in (and a subsidiary justification for) the construction of a proton or electron accelerator. Until recently, however, these schemes have rarely progressed beyond the conceptual stage. Table 3 lists details of the colliders in various stages of planning.

3.3.1. 30 GeV electron and 820 GeV proton collider – HERA (DESY, West Germany)

The West German national accelerator centre, DESY, has perhaps been the most active of the major high-energy physics laboratories in pressing for the construction of an electron–proton collider, and they appear to have been the first to prepare a detailed proposal (cf. Robinson [36, p.530]). Initially, the planned project (PROPER) involved adding a proton synchrotron to the existing electron–positron ring (PETRA) to yield collisions between beams of electrons at 20 GeV and protons at 280 GeV (cf. [5, p.364]). This scheme attracted strong support from the European Committee for Future Accelerators, who argued that it should be Europe's main priority after LEP. By 1980, however, DESY – at the time flushed with success as a result of completing PETRA two years ahead of a similar collider then being built by their great rivals at Stanford – adopted a more ambitious project for colliding 30 GeV electrons with 820 GeV protons. It was estimated that HERA, as it was named, would cost about 650 million Deutschmarks (some $265 million or 530 MSF) over a construction period of seven years (cf. Robinson [36, p.530]).[17]

The year 1980 also saw the German Federal Ministry for Research and Technology (BMFT) establishing an Advisory Committee (chaired by K. Pinkau to review and prioritize various Big Science projects then under consideration in the country. The Pinkau Committee, which reported in spring 1981, recommended that, in high-energy physics, continued German participation in CERN, and hence in LEP, be given first priority. In addition, it recommended approval in principle of the HERA proposal, but subject to certain reservations – in particular, urging that the initial electron–positron option be dropped, thus going some way towards assuaging the fears of those who believed that the energy of the electron ring might

Table 3
Major electron–proton (ep) colliders – present and proposed

	Date of first experiments	Maximum beam-energy (GeV)	Centre-of-mass energy (GeV)	Luminosity ($cm^{-2} s^{-1}$)	Estimated cost (MSF [a])	Number of experimental areas	Period of world leadership (years)	Future options
DESY HERA	1990	30 + 820	315	3×10^{31}	~ 540 [b]	4	5 + ?	?
KEK TRISTAN	~ 1990?	25 + 300	175	?	?	4?	0?	?
CERN ep [c]	mid-1990s?	130 + 270		?	?	?	?	?

[a] See note b to table 1.

[b] This is based on the most recent estimate of 660 million Deutschmarks at 1981 prices (cf. [20, p.90]).

[c] This project might involve colliding 130 GeV electrons from LEP with 270 GeV protons from the SPS, one of the possible long-term options for CERN.

---

[17] The electron ring would be built first, giving an option for carrying out 30 GeV electron–positron experiments in about 1986. This gave rise to fears among certain LEP supporters that the energy of this electron–positron collider might be "stretched" using superconducting radio-frequency cavities to achieve 50 GeV, and hence to reach the Z_0 region a year or so ahead of LEP.

be stretched to investigate the detailed properties of the Z_0 particle ahead of LEP (see footnote 17). The other main condition was that construction of HERA should not commence until 1984, partly for financial reasons (the limited funds available should first be used to exploit the existing collider, PETRA), and partly to give time for the further research and development work on the superconducting magnets required for the proton ring.[18] In 1983, BMFT decided to include in its planned budget for 1984 funds to enable construction of HERA to begin, provided that sufficient foreign participation in the project could be attracted (cf. [20, p.90]). It was estimated that construction would take seven years.

3.3.2. 30 GeV electron and 300 GeV proton collider – TRISTAN (KEK, Japan)

As we shall see below, the Japanese National Laboratory for High-Energy Physics (KEK) has embarked upon the construction of a 30 GeV electron–positron collider, TRISTAN. In a second stage of the project, it is planned to add a ring of 4.5 tesla superconducting magnets capable of accelerating protons to an energy of 300 GeV. Collisions of the proton beam with 25 GeV electrons will yield a centre-of-mass energy of about 175 GeV. If authorized in 1984, this project should be completed by about 1990 (cf. Trilling [40, p. 19]).

3.3.3. Other possible electron–proton colliders

As in the past, many of the new accelerators now being proposed or built envisage some form of future electron–proton option. For example, the Dedicated Collider project proposed by Fermilab (see above) incorporated an option for colliding 10 or 20 GeV electrons with 2 TeV protons (cf. Wojcicki [44, p.33]). CERN and Serpukhov also both have a future electron–proton option under consideration. At the former, electrons from LEP could be collided with protons from the SPS (cf. [8, p.5]). However, since it is now more difficult to obtain authorization for increasing the CERN annual budget (in view of the pressures on scientific funding in most Member States), the construction of such a facility could probably not be undertaken until the completion of LEP (and the achievement of a beam-energy of 130 GeV) in the early 1990s. Similar financial considerations suggest that Serpukhov's preliminary plans for colliding 20 GeV electrons with 3 TeV protons could also not be realized until well into the 1990s.

3.4. Electron–positron colliders

Because of the very great interest in ascertaining whether the Z_0 actually existed, there was tremendous eagerness in the high-energy physics community during the late 1970s and early 1980s to reach this energy region as quickly as possible. Although the CERN proton–antiproton collider achieved this goal first, its relatively low luminosity means that it will be able to probe only a small part of the physics of the Z_0. Hence a great deal of effort has been put into planning the construction of electron–positron colliders, both at CERN with LEP, and at a number of other laboratories. Table 4 lists the main details of LEP and its likely competitors.

3.4.1. 30 GeV TRISTAN (KEK, Japan)

In 1981, the Japanese Government approved the construction of the first phase of TRISTAN, an electron–positron collider capable of reaching a centre-of-mass energy of 60 GeV. It was estimated that the project would take five years to complete and cost a total of 7.5×10^{10} yen (about $350 million or 700 million Swiss francs) (cf. [10, p.103]). The collider is being built at KEK, the Japanese National Laboratory for High-Energy Physics, in the "Science City" of Tsukaba. KEK, like DESY in West Germany, is eagerly seeking active participation by overseas groups in the experimental exploitation of the machine (cf. [18, p.3]).

TRISTAN will operate initially with conventional radio-frequency cavities, but superconducting cavities are being developed at KEK and elsewhere which could permit the centre-of-mass energy to be raised at a later stage to approximately 90 GeV, bringing Z_0 physics within range (cf. [18]). In addition, as we saw earlier, there are plans to add a proton ring, thereby generating electron–proton collisions.

3.4.2. 50 GeV Stanford Linear Collider – SLC (SLAC, USA)

In 1977, when the European Committee for

---

[18] Brookhaven's traumatic experiences with ISABELLE showed all too clearly the dangers of proceeding with the construction of a new accelerator before all the necessary R&D on superconducting magnets had been completed.

Table 4
Major electron–positron colliders – present and proposed

	Date of first experiments	Maximum beam energy (GeV)	Centre-of-mass energy (GeV)	Luminosity ($cm^{-2}s^{-1}$)	Estimated cost (MSF [a])	No. of experimental areas	Period of world leadership (years)	Future options
DESY PETRA	1978	19 [b]	38	10^{32} [c]	–	4	8	–
SLAC PEP	1980	18	36	10^{32} [c]	–	6	0	–
KEK TRISTAN	1986	30	60	~ 10^{31}	~ 700	4	1	TRISTAN-45, ep
SLAC Linear Collider (SLC)	1987	50	100	6×10^{30}	240 [d]	1	1?	SLC-70
Cornell CESR-II [g]	1988	50	100	3×10^{31}	680	4	[illegible]	–
CERN LEP-50	1988	50	100	3×10^{31}	~ 1200 [e]	4	~ 2?	LEP-90/130, p, p̄p, ep
CERN LEP-90	~ 1990?	90	180	10^{32}	~ 350?	4	~ 2?	LEP-130, p, p̄p, ep
CERN LEP-130	early 1990s?	130	260	?	~ 400? [f]	8	3–5?	p, p̄p, ep
Novosibirsk VLEPP	mid-1990s?	150	300	~ 10^{32}	?	5	3–5?	VLEPP-500
SLAC Large Linear Collider?	mid/late 1990s?	1000	2000	~ 10^{33}	~ 6000	6	?	?

[a] See note b to table 1.
[b] Upgraded to 20 GeV in 1982, and 22.5 GeV in 1983.
[c] Yet to achieve the design figure.
[d] An additional 100 MSF will be needed to provide a second interaction region and new detectors (cf. Trilling [40, p. 52]).
[e] The collider itself will cost 910 MSF (in 1981 prices) and the four detectors a further 250–280 MSF.
[f] This excludes the cost of the four additional detectors (perhaps another 250 MSF).
[g] Project discontinued in 1983.

Future Accelerators decided that LEP should be CERN's next major facility, it could give as one of the justifications for this choice the argument that LEP would be unique, and therefore complementary to other new accelerators planned around the world. Now, however, this argument is less valid in view of the 50 GeV Linear Collider currently under construction at the Stanford Linear Accelerator Center (SLAC). In 1979, B. Richter began to explore the possibilities of modifying the laboratory's linear accelerator, upgraded in energy from the original 20 GeV to 50 GeV,[19] to collide electrons and positrons. Although technically extremely difficult (the principal problems is in developing accurate beams to ensure sufficient collisions), Richter was confident that this novel accelerator technology would provide a quick and cheap way of reaching the potentially important 100 GeV centre-of-mass energy region. The cost of the SLC was estimated in 1981 to be only $63 million [11, p.146]. Although revised to about $120 million in 1982 (cf. Trilling [40, p.52]), this is still only about one quarter of that being spent on LEP to achieve the same energy (although LEP can be subsequently extended to attain much higher energies). Stanford did not obtain formal governmental authorization for the project to begin until Fiscal Year 1984 (cf. [19, p.81], but it had been funding the necessary R&D since 1980 (using the laboratory's operating budget), with actual construction beginning in 1981 (cf. [15, p.8]). It was estimated in 1983 that the first collisions will be obtained by early 1987 (cf. Wojcicki [44, p.25]), just over a year ahead of LEP.

3.4.3. 50 GeV CESR-II (Cornell, USA)

Encouraged by their success in the late 1970s with the CESR electron–positron collider (see Paper II [24]), Cornell scientists began to press their case to enter the race to discover the Z_0 particle, proposing the construction of a second-generation electron–positron storage-ring (CESR-

---

[19] If there was sufficient physics interest, the energy of the SLC could subsequently be increased to 70 GeV (this is referred to below as SLC-70).

II). By using superconducting radio-frequency cavities (see footnote 20 below), it was estimated that a beam-energy of 50 GeV might be reached in early 1988, some months ahead of LEP, and for just over half the cost ($340 million – cf. Trilling [40, p.40]). However, the scientific arguments advanced for CESR-II were virtually identical to those for the Stanford Linear Collider; and, in the opinion of the large majority of scientists interviewed in the course of our study, the latter had at least three major advantages: it was approximately a third the price of CESR-II, it provided a test-bed for the new technology of colliding linear accelerators – a development of great potential importance for the future of accelerator design (see below); and CESR-II required a new and thus far unproven technology (superconducting cavities) that may not yet be suitable for mass production.[20] In view of this last point, Cornell made it clear in 1982 that they would not submit a firm proposal until they had satisfactorily demonstrated the operation of superconducting radio-frequency cavities and that mass-production techniques for their fabrication had been established (cf. [40, p.39]). At that stage, Cornell expected to be in a position by 1984 to submit a proposal for commencing construction of CESR-II in Fiscal Year 1985. However, given the straitened financial circumstances confronting US high-energy physics, it was decided in 1983 not to proceed with CESR-II. Instead, funds were provided to upgrade the existing collider (in particular, to improve its luminosity) and to continue R&D on superconducting cavities for possible use in future machines.

4.4. 150–500 GeV VLEPP (Novosibirsk, USSR)

The Soviet high-energy physics centre at Novosibirsk in Siberia has, as we have noted elsewhere (Irvine and Martin [25]), developed into a world centre of excellence for electron–positron accelerator technology, even though the actual research output from the laboratory has not been significant in international terms. Their latest proposal is to construct a machine capable of colliding beams, each initially of 150 GeV energy, from two linear accelerators. At a later stage, the length of the linear accelerators could be increased to give beam-energies of up to 500 GeV (cf. [17, pp.417–18]). However, in view of the Soviet Union's existing commitment to the construction of UNK, it would seen doubtful whether substantial funding for VLEPP could be made available until the late 1980s.[21] Consequently, even if the project is eventually authorized, it is unlikely to be completed before the mid-1990s.

3.4.5. 1000 GeV Large Linear Collider (SLAC, USA)

Although the Stanford Linear Collider (see above) is still some way from completion, staff at SLAC have already begun to consider a scheme to extend current linear-accelerator technology to the 1000 GeV (1 TeV) energy range. No definite proposal has yet been prepared, but such a project was put before the US Subpanel on New Facilities in 1983. It was estimated that beam-energies of 1 TeV could be achieved for a cost of about $3 billion (cf. Wojcicki [44, p.37]). Given SLAC's current commitment to the early completion of the SLC and the fact that the necessary preproposal R&D will take at least five years, such a facility could probably not be completed until the latter part of the 1990s.

4. A framework for comparing the future prospects of new experimental high-energy physics facilities

Having looked in some detail at the accelerators currently planned or under construction around the world, we are now in a position to begin identifying the comparative advantages and disadvantages of LEP and its likely competitors. As noted earlier, our aim has been to develop a framework for systematically assessing, from an outsider's perspective, the future prospects of capital-intensive projects in basic science. The case for such assessments in high-energy physics has perhaps been reinforced by the evidence presented in the previous section concerning not only the very substantial costs involved (and indeed the possibility of failure as in the case of ISABELLE), but also the somewhat *ad hoc* way in which such facilities have tended to be planned, authorized and constructed, something which must be a source of some concern given the trend towards increasingly powerful institutional pressure groups discussed at the start of the paper.

---

[20] It should be noted that CERN also considered employing superconducting radio-frequency cavities in the first phase of LEP, but decided against this because of the attendant technological risks: "Superconducting cavities are not yet sufficiently mastered. LEP Phase 1 must incorporate 'conventional' technology ..." [16, p.63].

[21] Strictly, the Novosibirsk Institute belongs to the USSR Academy of Sciences (Siberian Branch), while Serpukhov is an Institute of the State Committee for Atomic Energy, so VLEPP would be funded from a different source than UNK. Even so, it seems unlikely that the Soviet Union could afford to commit itself to providing the resources for two very large high-energy physics projects simultaneously.

In what follows, we apply a set of thirteen criteria to assess the future prospects for LEP in relation to those for other new experimental facilities. The criteria were identified on the basis of our studies of the factors determining the relative success and failure of high-energy physics accelerators (see Paper II [24]); they relate mainly to future scientific potential, but also take into account factors such as relative resource requirements and potential for the development of future research facilities. Table 5 evaluates the various new accelerators planned to come into operation over the next decade or so against these criteria. In addition, the recently completed proton–antiproton collider at CERN and the now discontinued ISABELLE and CESR-II projects have been included for comparative purposes. Each machine is assessed in terms of the various criteria on the basis of material and data discussed in the text below. As will be seen, this table provides a convenient means of summarizing the patterns of comparative advantage (indicated by + signs) and weakness (− signs) that exist among the various projects, patterns that can then be used in arriving at an overall assessment of their future prospects. We begin by reviewing the first set of four criteria relating to the relative construction requirements of these various accelerators.

4.1. Comparison of construction requirements

4.1.1. Financial criteria

Relative costs. In assessing the merits of different projects, relative cost is clearly one of the principal criteria. The estimated costs (in millions of Swiss francs or MSF) of the main accelerators and colliders under consideration were given in tables 1–4. Of these, the cheapest are the CERN proton–antiproton collider (200 MSF), the Stanford Linear Collider (240 MSF), and the Fermilab 1 TeV accelerator (about 250 MSF), their low costs reflecting the fact that they all make significant use of already existing facilities. In contrast, the 50 GeV first phase of LEP (910 MSF plus a further 250 to 280 MSF for four detectors) and ISABELLE (estimated to cost a total of some 1300 MSF at the time the project was discontinued, although there was no guarantee that this figure had ceased escalating) are both considerably more expensive. However, even their costs are dwarfed by that of the proposed US Superconducting Super Synchrotron and its associated detectors (some 6000–8000 MSF). These relative levels of cost are summarized in the first row of table 5.

Accessibility of resources. One of the major areas of uncertainty in assessing the future prospects for an accelerator is the likely funding position of rival facilities. This can determine not only whether competing accelerators are built, but also their date of completion – a factor which is often crucial in structuring success and failure. In the case of the first phase of TRISTAN, UNK, and HERA, the respective national governments have already authorized the necessary expenditure. For the Fermilab 1 TeV accelerator and collider, the Stanford Linear Collider, and the first phase of LEP,[22] the resources have been promised, but for all these projects there is little leeway for cost over-runs (apart from extending the construction period). These conclusions on the probable accessibility to financial resources are summarized in the second row of table 5.

4.1.2. Technical criteria

Degree of technical difficulty. As accelerator construction has increasingly become a sophisticated high-technology activity, so the technical difficulty of constructing and operating such a facility has become a more important factor in determining its likely future success. This is particularly true for very advanced technologies like superconductivity. Recognizing the difficulties involved, Fermilab adopted a relatively cautious approach to the design and construction of the superconducting magnets required for its 1 TeV accelerator. Even then, the task proved far from simple and it was only in 1980, after a long development process, that they were able to begin manufacturing mag-

---

[22] Once this first phase of LEP has been completed, CERN will in all likelihood be able to convince the Member States that rather than decreasing the annual budget to the "base-load level", any free funds should be used to increase LEP's energy to 90 GeV and later to 130 GeV.

Table 5
Comparison of major accelerator projects: Summary of main relative comparative advantages (+) and disadvantages (−)

			CERN $p\bar{p}$	Fermilab 1 TeV	Fermilab $p\bar{p}$	KEK TRISTAN	SLAC SLC-50	Cornell CESR-II [a]
		Estimated date of first experiments	1981	1985	1986	1986	1987	1988? [b]
Construction requirements	Financial	Cheapness	+ +	+ +	+	−	+ +	−
		Accessibility of resources	+ + +	+	+	+ +	+	− − −
	Technical	Technology required relatively simple/undemanding	− −	−	−	+	− −	−
		Technical track-record of laboratory	+ +	−	−	·	+	·
Scientific potential	World lead in a new energy region	Relative increase in centre-of-mass energy	+ + +	·	+ +	·	+	−
		Period of world leadership	+ +	+ +	+ + +	·	·	−
	Other factors governing ability to generate physics results	Event rate	− −	+ +	− −	·	−	·
		Number of experimental areas	−	+ +	−	·	− −	·
		Variety of experiments possible	·	+ +	·	−	− −	−
		"Cleanness" (ease of interpretation) of data	−	·	−	+ +	+ +	+ +
	Ability of users to exploit accelerators	Scientific track record of laboratory users	+	·	·	−	+ +	·
Future potential		Potential spin-off to accelerator physics	+	+	·	·	+ +	+
		Flexibility/potential for future development of accelerator	−	+	−	+	·	·

[a] Project discontinued in 1983.
[b] This assumes that a superconducting proton collider is not completed at CERN ahead of this US machine.
[c] '·' signifies no particular comparative advantage or disadvantage.

nets on a mass-production basis. At Brookhaven, where the ISABELLE project required magnets with significantly higher field-strengths, the problems were far more severe, and were largely responsible for the delays to the facility's construction programme,[23] and hence ultimately for the project's cancellation. The technical difficulties faced by the CERN accelerator builders in attempting to achieve a luminosity of 10^{30} cm^{-2} s^{-1} in the proton–antiproton collider have also been extreme, the machine initially operating with a peak luminosity a factor of 200 below the design figure, although this was rapidly improved by one and a half orders of magnitude over the following two years. The builders of the Fermilab collider will enjoy the considerable advantage of being able to benefit from the experiences at CERN, making their task markedly simpler. In the same way, Serpukhov in their collaborative work with Saclay on the superconducting magnets for UNK will be able to profit from Fermilab experience with the magnets for the 1 TeV fixed-target accelerator, as indeed will the builders of HERA.[24]

---

[23] As noted earlier, one of the factors contributing to this delay was the fact that the original design for the ISABELLE magnets had eventually to be scrapped after a long period of unsatisfactory development and testing. Part of the problem may have stemmed from the Brookhaven design being so different from that adopted at other high-energy physics centres that the laboratory was unable to profit from the results of work elsewhere (cf. [11, p.146]).

[24] The proposed magnet designs for both UNK and HERA are reportedly similar to that adopted at Fermilab (cf. [11, p.147] and [13, p.206]).

Table 5 (contd.)

			CERN LEP-50	Brookhaven ISABELLE[a]	Serpukhov UNK	DESY HERA	KEK ep	CERN LEP-90/130	US SSC
		Estimated date of first experiments	1988	1987/8?	1990?	1990?	1990?	early 1990s?	mid-1990s?
Construction requirements	Financial	Cheapness	− −	− −	+	[illegible]		−	− − −
		Accessibility of resources	+	− − −	+ +	+ +		+	
	Technical	Technology required relatively simple/undemanding	+ +	− −				+	−
		Technical track-record of laboratory	+ +	− −	− −			+ +	?
Scientific potential	World lead in a new energy region	Relative increase in centre-of-mass energy	−	− −	+	+ + +	− −	+	+ + +[a]
		Period of world leadership	+	− −	+ +	+ +	− −	+ +	+ +
	Other factors governing ability to generate physics results	Event rate		+	+ +			−	+
		Number of experimental areas			+ +			+	?
		Variety of experiments possible	−	+	+ +	+	+	−	+
		"Cleanness" (ease of interpretation) of data	+ +	−		+	+	+ +	−
	Ability of users to exploit accelerators	Scientific track record of laboratory users	+	+ +	− −		−	+	?
Future potential		Potential spin-off to accelerator physics		+				+	+
		Flexibility/potential for future development of accelerator	+ +		+			+	?

In the case of electron–positron colliders, the technical problems associated with the SLC are felt by most physicists to be particularly severe, and it will be a major engineering feat to achieve a luminosity of 6×10^{30} cm^{-2} s^{-1} with the new and untested technique of the linear collider. In contrast, the task facing the builders of TRISTAN and LEP seems more simple. For the latter, one of the main problems – that of producing low-field magnets relatively cheaply – has already been largely solved (cf. [8, pp. 8–9]), and the difficulties anticipated in tunnelling under the nearby Jura mountains have also been considerably diminished by various changes in the size and location of the accelerator ring. Finally, although there may be inherent problems in reaching the design luminosity with a circular collider like LEP.[25] CERN will be able to profit from the wealth of experience gained over a ten-year period from smaller elec-

---

[25] In 1981, it was reported that "CESR, PEP and PETRA are all having great difficulty climbing anywhere near their design luminosity" [11, p.144]. One commentator concluded somewhat pessimistically that this "low luminosity also cast a shadow on the performance prospects of larger colliding beam machines still on the drawing board" (Robinson [33, p.1488]). There are perhaps two underlying causes to the disappointing results so far. One is that calculations of the design luminosity for the newly completed colliders were based on extrapolations from a previous generation of machines, and these extrapolations have subsequently proved to be invalid. (It is therefore somewhat alarming that "Published CERN plans for [LEP] still use the old tune-shift value in calculating the expected luminosity" [33, p.1490] rather than the values actually obtained at PETRA and PEP.) The other factor is that, in the rush to produce high-energy physics results as quickly as possible, those responsible for the operation and development of electron–positron colliders have given too little attention to the machine physics involved (cf. [11, p.144]).

BEST COPY AVAILABLE

tron–positron colliders at DESY, Cornell, and Stanford. (If and when LEP's energy is increased to 130 GeV, superconducting cavities will be required, but by then the technical uncertainty associated with their production and use should be fairly modest, and CERN will be able to draw on the many years of work currently being undertaken at Cornell, Karlsruhe, KEK, and elsewhere.) These comments on the relative technical difficulty associated with the various accelerator projects are again reflected in the relevant row of table 5.

Technical track-record of the laboratory. One of the best indicators of the likelihood of overcoming the technical diffculties associated with designing, building, and operating a new facility is the previous record of the accelerator-builders concerned. CERN has perhaps had the best technical record of the major accelerator centres over the last decade with the successful commissioning of the ISR [26] and the SPS,[27] a fact widely acknowledged by the physicists we interviewed in both Europe and the USA. In view of this, it was perhaps not unreasonable to assume when CERN embarked upon the construction of the proton–antiproton collider that the very considerable technical problems would eventually be solved, and this assumption has been fully justified by subsequent events. Similarly, the technology required for LEP should prove easily within CERN's competence, although, if the tunnelling encounters any major difficulties, the tightness of the budget may result in the date of the first experiments being postponed.

SLAC also possesses an excellent track-record with the linear accelerator and particularly with SPEAR, which was built under very adverse financial constraints (funds had to be found from within the laboratory's existing budget after the project application had been turned down by the US Department of Energy). The more recent experience with PEP, however, was markedly less successful, principally because of delays caused by outside civil-engineering contractors, a problem that CERN may face in connection with the LEP tunnel. However, on balance, SLAC's record suggests that it will probably be able to solve the problems associated with the SLC and complete the facility by 1987 as planned. In recent years, DESY has had almost as strong a technical track-record as SLAC. It has grown into a major international laboratory, and with this has come the breadth of technical and engineering expertise that will probably ensure it is able to cope successfully with the demands of HERA even though this project is rather more ambitious than previous construction programmes. Fermilab, in contrast, has a somewhat patchy technical record with its 400 GeV accelerator, as was seen in Paper II [24]. Nevertheless, there are signs that the laboratory has since taken note of the lessons learnt, particularly in magnet construction, and neither the 1 TeV accelerator nor the proton–antiproton collider appear beyond its technical capabilities.

One cannot be quite so sanguine about the prospects for UNK. As we have discussed elsewhere (Irvine and Martin [25]), Serpukhov does not have a particularly strong record for building front-line research facilities, probably for reasons more to do with the structure of East European science than with the ability of individual accelerator physicists. Although the 70 GeV accelerator first operated in 1967, it was several years before a full experimental programme was mounted, and, even then great difficulty was encountered in providing the sophisticated detectors and powerful computing facilities needed to progress beyond the simplest "first generation" experiments. Perhaps because of the limited access of East European high-energy physicists to the services of high-technology industry, it has been decided that the superconducting magnets for UNK will be built at Serpukhov itself (cf. [11, p.147]). While this may give the laboratory greater control over the construction of the accelerator than might otherwise have been the case, it is still difficult to be optimistic that all the problems encountered in Western attempts to mass-produce such magnets will be quickly and efficiently overcome by the Soviet team, even with the aid of their French collaborators from Saclay.

Similar reservations applied to Brookhaven with respect to the ISABELLE project while it was still proceeding. Although Brookhaven had a good record with the AGS two decades ago, the AGS

---

[26] See the quotation in section 8 of Paper II [24] on the ISR being "widely regarded as the most perfect example to date [1979] of the accelerator builder's art."

[27] In Paper II [24], it was seen how the SPS's technical superiority over the Fermilab accelerator was one of the main factors explaining the differences between the scientific performance of the two machines in the late 1970s and early 1980s.

Conversion Programme undertaken ten years later experienced severe problems (see Paper II [24]), and the record of the laboratory with developing superconducting magnets between 1976 and 1980 can only be described as poor. As has been pointed out, when the proposal for ISABELLE was being formulated, the machine had no obvious

> competitor in its particular regime of physics, and [the project] had time for R&D – more time, due to stringent funding of high-energy physics, than [Brookhaven] was happy with (Metz [31, p.188]).

There must therefore be considerable doubts about the wisdom of the laboratory in embarking upon the construction of the new facility before all the necessary R&D work had been completed,[28] and in continuing a commitment to an inadequate magnet design for several years after its limitations became apparent to those inside the laboratory. Although changes in the senior management at Brookhaven during 1981 and 1982 led to some technological revitalization of the project,[29] they did not completely remove what was probably the root cause of Brookhaven's relatively poor technical record during the 1970s – the failure to recruit capable young staff, especially accelerator physicists and engineers (cf. [6, p.247]) – which was in turn brought about by the absence over a longer period of any major new accelerator project.

4.2. Comparison of scientific potential

The scientific justifications made for all the new high-energy facilities of the 1980s are remarkably similar, and exhibit a degree of convergence of theory and experimental practice seldom previously witnessed in the subject. Analysis of the proposals for each of the accelerators listed in table 5 (and of subsequent review papers and reports summarizing the scientific case for each facility) reveals a common emphasis on the study of the intermediate vector bosons, the search for the Higgs particle and for new leptons and quarks, investigation of the quark–gluon picture and the theory of quantum chromodynamics, the exploration of possible "grand unified theories", and the discovery of unexpected phenomena. How, then, can one attempt to draw distinctions between the scientific potentials of these various accelerators? What criteria can be used?

The history of experimental high-energy physics over the last thirty years shows that two of the main factors determining the scientific potential of an accelerator are whether it opens up a new, unexplored energy-region (i.e. whether it has a significant advantage in terms of increased centre-of-mass energy over other machines), and the length of time it enjoys a position of world leadership for that particular type of accelerator. As was seen in Paper II [24], this was certainly a crucial component in the success of proton machines such as the Berkeley Bevatron, the Brookhaven AGS (with its small but significant advantage in energy over the CERN PS), and the Fermilab accelerator. This pattern is also reflected both in the case of electron machines, notably in the output of the Stanford linear accelerator, and among electron-positron colliders with the success of first SPEAR and then PETRA (see footnote 33 below).

A second set of factors, and one which has become more apparent in recent years, relates to the varying potentials of different types of accelerator for generating new physics results. Several factors are of relevance here: the event-rate (which determines the amount of experimental data that can be collected in a given period), the number of experimental areas or interaction regions (which determines the number of experiments that can be undertaken at any one time), the variety of experiments possible on the accelerator, and the degree of "cleanness" of the experimental data produced (i.e. the ease with which they can be interpreted).

Last, and of great importance, is the ability of the user-group associated with the accelerator to exploit its full potential. As we have argued elsewhere (e.g. Martin and Irvine [28]), one of the few indicators of that ability is their recent track-record, and therefore this criterion has also been included in table 5.

Let us examine in turn each of these criteria relating to the scientific potential of the various accelerators under consideration.

---

[28] In the attitude survey described in Paper II [24], 74 percent of those interviewed agreed with the statement, "It would have been better if Brookhaven had first spent a few years ascertaining whether superconducting magnets were technologically feasible before investing considerable resources in the construction of ISABELLE", well over three times the number who disagreed (22 percent).

[29] Prototype magnets of the necessary field-strength were successfully developed in 1982, although at the time the project was cancelled it still had to be demonstrated that these were suitable for mass-production (cf. Wojcicki [44, p.30]).

4.2.1. World-lead in a new energy region

Relative increase in centre-of-mass energy. Taking first the proton machines, one can see from table 2 that the centre-of-mass energy achievable with the CERN proton–antiproton collider is 540 GeV. This represents almost a nine-fold increase over that attained on the previous highest-energy proton collider, the CERN ISR – an unusually large jump in energy that has, in the past, not been surpassed by any other accelerator (the ISR came closest with a five-fold increase in centre-of-mass energy over Serpukhov). As for the rival Fermilab proton–antiproton collider, its energy advantage over the CERN facility will correspond to a factor of just over three and a half (2000 GeV compared with 540 GeV), again quite large in terms of previous experience. In contrast, the Fermilab 1 TeV accelerator will reach a centre-of-mass energy less than 50 percent higher than the maximum achieved by the current generation of large proton synchrotrons, while UNK will in turn eventually possess an advantage of about 75 percent over the Fermilab accelerator. Finally, it is worth commenting that ISABELLE, if completed, would have had a centre-of-mass energy very much lower than the collider at Fermilab, although (as is noted below) it would have had the advantage of a higher event-rate.

As for the new electron machines, while TRISTAN's energy will be only 20 percent greater than that of the upgraded PETRA (see table 4, note b), the Stanford Linear Collider will achieve an energy 70 percent greater than TRISTAN – an increase which, as we have seen, could be crucial because it will permit detailed studies to be carried out of the mass, width, and decay-modes of the Z_0 particle. If the SLC is completed in 1987 as planned, then LEP will not have any energy advantage when it begins operating a year later, though it will have other relative strengths (see below). However, if the beam-energy of LEP is eventually increased to 130 GeV, this will give the machine a two-fold energy advantage over other electron–positron colliders – in particular the SLC, even if the latter's energy is subsequently increased to 70 GeV. Finally, the DESY electron–proton collider, HERA, will achieve a centre-of-mass energy over forty times greater than previously achieved in electron–proton collisions – 315 GeV compared with less than 7 GeV at the Stanford linear accelerator – so, like the CERN proton–antiproton collider, it is highly ranked in terms of this criterion in table 5.

Period of world leadership. The penultimate columns of tables 1–4 contain estimates of the number of years each accelerator is likely to produce the highest-energy interactions for that particular type of facility. (It should perhaps be stressed that, particularly for those machines still in the planning stage, these are very approximate estimates only.) The relevant row in table 5 shows which facilities are likely to enjoy a relatively long period of world-leadership, and which are not. Thus the rating reflects the fact that, for example, the CERN proton-antiproton collider is likely to enjoy a lead of some five years over its Fermilab rival, while it may be up to ten years before the latter is overtaken by a yet higher-energy proton collider (perhaps one placed alongside the electron–positron ring in the LEP tunnel or the SSC). The Fermilab 1 TeV accelerator, UNK, HERA, and the final phase of LEP (LEP-130), are all given reasonably positive evaluations due to the fact that they should enjoy a world lead of about five years (± two years). This is considerably more than the TRISTAN and SLC machines, both of which are likely to be overtaken by more powerful electron–positron colliders within a year or so. Finally, it should be stressed that the ill-fated ISABELLE project again ranked relatively low in terms of this criterion because it would have come into operation with a marked energy disadvantage relative to other machines as a result of the delays to the construction programme discussed earlier.

4.2.2. Other factors governing ability of accelerator to generate new physics results

Event-rate. The rankings of the various accelerators listed in table 5 in terms of this criterion reflect substantial differences among accelerators in the event-rates they are able to generate. The fixed-target machines (such as the Fermilab 1 TeV accelerator and UNK) have by far the highest event-rates, followed by proton–proton colliders (the proposed SSC is likely to have a design luminosity of approximately 10^{33} cm^{-2} s^{-1}) and electron–positron storage-rings like TRISTAN and LEP (with design luminosities in the region of

10^{31} – 10^{32} cm^{-2} s^{-1}).[30] The event-rates of the proton–antiproton colliders at CERN and Fermilab are appreciably lower (of the order of 10^{30} cm^{-2} s^{-1}), with the result that experiments carried out on these machines take correspondingly longer, and consequently there are fewer of them. Conversely, proton machines can be used to make relatively quick scans of large energy regions.

Number of experimental areas. Fixed-target accelerators like the Fermilab 1 TeV and UNK also possess a marked advantage over other sorts of experimental facility in that several beams can be obtained simultaneously. As a result, large numbers of experiments can be run concurrently – far more than are possible on a collider where the number of interaction areas is strictly limited. In the final phase of LEP, for example, eight experimental areas are planned, while in the first phase (LEP-50), there will be only four. At the CERN and Fermilab proton–antiproton colliders, the number of experiments is even more constrained, with just two intersection regions, while at the Stanford Linear Collider there will probably be only one collision point, at least initially.[31]

Variety of experiments. As is highlighted in table 5, fixed-target accelerators possess yet another important advantage over colliders; because they can be used to produce a wide range of secondary beams (of kaons, neutrinos, etc.), the variety of different types of experiments – or what is often termed the "physics menu" – is much broader than for electron–positron colliders (and especially so in the case of the SLC with its relatively low luminosity[32]). This has a number of important policy implications which should perhaps be briefly mentioned at this point, though they are treated in more detail in the conclusions. The first is that, with electron–positron colliders, there is a crucial advantage to be gained by being first into a new energy region;[33] a lead of just one or two years for such a machine can be more important than a head-start of three or four years for a fixed-target accelerator where it takes considerably longer to exhaust the range of experimental possibilities. Hence, the fact that the SLC is due to reach the 100 GeV centre-of-mass region a full year or so ahead of LEP could prove a major handicap during this first phase of the CERN machine. A second policy implication has been described in the following terms in relation to LEP:

> There is a natural tendency when investigating a new area of physics to be the "firstest with the mostest." Since it is possible to define all the characteristics we would like to know about the products of electron–positron collisions, this tendency could lead to all experimental proposals being built around nearly identical "universal" detectors. [14, p.241]

However, CERN management seem to be aware that such a tendency, if not checked, could leave LEP particularly vulnerable to a sudden dramatic change in physics interest, since they have urged that the differences between the various detectors proposed for LEP be made rather greater than first seemed likely (cf. Robinson [38, p.722]). Even so, the variety of experiments that can be handled on an electron–positron machine like LEP is appreciably less than that possible on some of the other new accelerators planned to come into operation in the next ten years.

"Cleanness" of data. Where electron–positron colliders do score highly over other types of accelerator is in terms of the ease with which the resulting experimental data can be interpreted, this facet of their performance being strongly reflected in table 5. The reason for this is as follows: collisions between electrons and positrons involve an interaction between two apparently point-like objects, each with precisely known energies. When these particles collide, total annihilation takes place, producing "pure" energy which is then available for the creation of new particles that have a known

---

[30] In the case of LEP-90, the design luminosity is 10^{32} cm^{-2} s^{-1}. The luminosity in the earlier phase (LEP-50) is expected to be 3×10^{31} cm^{-2} s^{-1}, giving it a probable advantage in this respect of about 5 over the similar energy Stanford Linear Collider (SLC-50).

[31] It is envisaged that this may subsequently be increased to two.

[32] The SLC will, however, have the compensating advantage that it is uniquely suitable for polarized beam experiments (cf. [12, p.205]).

[33] This was particularly true in respect of the Stanford collider SPEAR (a year ahead of DORIS at DESY), and later, when the positions were reversed, with PETRA coming into operation two years ahead of PEP. The rapid obsolescence of electron and electron–positron machines is further discussed in Martin and Irvine [27].

total energy. Hence, by successively "tuning" the energy of the collided electrons and positrons to a series of different values, a particular energy region can be comprehensively scanned – if any new particle does exist in the energy-range under study, then it should be produced and thus be observed.[34] In contrast, the collision between two protons (or between a proton and antiproton) is a much more complex affair,[35] since each particle (according to the currently accepted "standard model") consists of three quarks and a number of gluons. In the resulting multi-body interaction, only a fraction of the collision energy is available for the production of new particles; and since the energies of the individual constituent quarks are unknown, it is impossible to "tune" the experiment to investigate a particular energy state. Instead, a whole spectrum of particle interactions of different energies takes place, and the experimentalist has to sift painstakingly through all of these to find the collisions of interest; in other words, to use the jargon of physicists, the events are "dirty" (cf. Robinson [35, p.528]). Hence, despite the development by high-energy physicists of sophisticated computer programs to perform such analyses, and the enormous growth in recent years of available computing power, one can never be entirely sure that nothing has been "missed" in a given energy region.[36] Last, HERA lies somewhere in between electron–positron colliders on the one hand, and proton machines on the other, since it involves using electrons – i.e. apparently point-like objects with reasonably well known properties – as a probe to investigate the structure and properties of the multi-body proton.

4.2.3. Ability of users to exploit accelerator

As noted earlier, the track-record of the user-community associated with a laboratory is a crucial variable in determining whether the scientific potential of a new accelerator is likely to be achieved. The data presented on this subject in table 5 reflect the differences in scientific performance identified in Paper II [24] for the world's principal accelerators over the last two decades. Thus, particularly high credit must be given to the user-communities associated with Brookhaven (the AGS) and Stanford (the linear accelerator and SPEAR), which were judged to have the most successful records in world terms – each was responsible for a comparatively large number of crucial discoveries, as well as a significant volume of lower-level advances. In addition, the successes of the proton–antiproton collider in 1983 have clearly strengthened the record of CERN users. At the other end of the scale, Serpukhov had the least distinguished record of all the major laboratories currently engaged in a major new accelerator project (see Irvine and Martin [25]). However, these data on past performance cannot legitimately be used in prospective evaluation without a number of qualifications first being stated. In the case of Brookhaven, for example, it should be noted that the main achievements of the AGS were made in the 1960s by a group of predominantly US East Coast experimenters, while the user-community for ISABELLE (if it had finally been completed in 1987 or 1988) would have been very different. A similar consideration may apply to Stanford, though probably not to the same extent. In the heyday of SPEAR, most of its principal users were employed either at the laboratory itself or at the Lawrence Berkeley Laboratory. Since then, the Stanford user-community has been considerably broadened,[37] a trend that seems likely to continue. In general, the future will probably witness even greater cross-usage of accelerators than in the past because the growing cost of new experimental facilities means that there will be less duplication of accelerators between continents. So, for example, large numbers of European physicists (perhaps 400 or so – see Robinson, [39, p.814]) are set to use the new Fermilab machines, especially the proton–antiproton collider,[38] once these come into operation as the highest-energy proton facilities in the world. Similarly, when LEP takes over the mantle as a world's most powerful electron–positron storage-ring, there will be a major movement of American experimentalists to CERN,[39] while LEP and HERA will be drawing users from essentially the same community of West European experimentalists (which may give rise to an element of competitio[illegible]een them). Such migrations of physicists between centres are, therefore, likely to go at least some way towards smoothing out the differences between the ability of traditional user-groups to exploit their experimental facilities, although the users of LEP will still be predominantly West European (perhaps 75 percent), those of the Stanford Linear Collider mainly West Coast Americans, those of UNK primarily East European, and so on. Therefore, track-record is still likely to constitute an important factor in predicting future research performance.

---

[34] This largely explains the tremendous success of SPEAR in the mid-1970s in exploring the range of "charmonium" particles, and, more recently, of CESR (at Cornell) in investigating the upsilon and related particles.

[35] In Paper II [24], we saw that this was one of the factors limiting the relative scientific performance of the CERN ISR.

[36] It is precisely because of the enormous problems with data-handling in proton–proton physics that critics of ISABELLE argued that its luminosity – the only remaining advantage it had over rival facilities – was only of marginal importance.

[37] This was a strategic decision taken in the latter part of the 1970s by the SLAC management, who became increasingly concerned that, in a period of limited funds, the scientific case for the laboratory's future support might be swamped by the political strength of the wider body of university experimentalists whose research activities were concentrated around the more extensive fixed-target programmes of the two national laboratories at Brookhaven and Fermilab.

4.3. Comparison of future potential

Finally, we come to the last set of factors listed in table 5 that should be taken into account in assessing the future prospects of a major accelerator project. These concern, first, the likely spin-off from a new machine to accelerator physics more generally – for example, whether a radically innovative technique for collisions is being pioneered that may usher in a new generation of research facilities; and, second, the "flexibility" of the machine – that is, its potential for future development (such as adding a new ring to generate further experimental possibilities) or for deployment in other scientific specialties. Some of the physicists interviewed particularly stressed these factors, arguing that, if a particular new energy-region should prove barren, it is important to be able to make alternative use of what are highly expensive capital facilities.

Potential spin-off to accelerator physics. Of the various accelerator projects considered here, the Stanford Linear Collider probably has the greatest potential in terms of spin-off, since it provides an opportunity for testing the feasibility of colliding linear accelerators. This new accelerator technology is important because it now seems likely that LEP will be the last of the large circular electron–positron colliders. Not only would a larger collider than LEP be prohibitively expensive to construct (the cost of such machines rises approximately with the square of its energy), but the power consumption would be extremely high[40] (to overcome the energy losses through synchrotron radiation). Furthermore, the beams would be subject to a phenomenon know as beamstrahlung,[41] which would seriously limit the accelerator's performance. To achieve electron–positron collisions at centre-of-mass energies greater than 350 GeV or so, the most promising way forward would appear to lie with colliding linear accelerators (the cost of which rises only linearly with energy). According to many of the physicists we interviewed, the SLC will provide an important test of whether this new and potentially important technique is feasible,[42] and it is therefore rated highly in terms of this criterion in table 5.

As for the other new machines, the CERN proton–antiproton collider has been technologically very innovative in that it has involved developing the technique of "stochastic cooling", which has the potential for extensive future applications. The main spin-off from the Fermilab 1 TeV accelerator (and from the ISABELLE project while it was still being funded) has been the technology and expertise required in the mass-production of superconducting magnets. Finally, the last phase of LEP (LEP-130)[43] requires the development of superconducting radio-frequency cavities, as would CESR-II had it been built. However, in none of these cases is the magnitude of the potential spin-off as great as for the Stanford Linear Collider.

---

[38] There has been some discussion, for example, about the possibility of moving the very large UA1 detector from the CERN collider to the larger US facility when it nears completion in 1985.

[39] Of the four LEP collaborations given the provisional go-ahead in 1982 by CERN, one is largely American (cf. Robinson [38, p.722]).

[40] Indeed, it will take over 250 MW of electricity to power LEP at 90 GeV, the equivalent energy consumption of a medium-sized city. It was for this reason that a senior American accelerator physicist intimately involved with the development of the SLC referred to LEP in an interview as "the last of the dinosaurs."

[41] This is the "radiation" phenomenon arising from high-field effects as bunches of particles pass through one another at very high energies.

[42] It is significant that the SLC is categorized as a "research and development project," although it will have "a timely physics payoff as well" (Wojcicki [44, p.25]).

Flexibility. The variety of possible options for future experimental facilities is probably greatest in the case of the first phase of LEP (LEP-50). As we saw earlier, once the later phases of the accelerator have been completed (LEP-90 and LEP-130), the possibility exists for placing a proton synchrotron in the same tunnel.[44] (The width of the LEP tunnel has been made sufficiently large to cater for such future developments, thus minimizing the civil engineering costs associated with any major new project of this sort.) As we have already noted, by using superconducting magnets, a proton synchrotron with an energy of up to 10 TeV[45] could be built on the CERN site. This accelerator could in turn be converted into a large proton–antiproton collider and be used, in conjunction with LEP, to collide 130 GeV electrons and 10 TeV protons. This potential for future development was referred to by many of the CERN physicists interviewed in 1981 as the "real-estate" argument for LEP – that, regardless of the scientific potential of the electron–positron collider, the existence of the costly 27-kilometre tunnel will guarantee the long-term scientific future for CERN (and West European high-energy physics) by virtue of the wide range of new accelerator developments that can be accommodated at the Geneva site. However, it is significant that, since these interviews were conducted, US high-energy physicists have put forward plans for the SSC (a 20 + 20 TeV proton collider). In 1983, the High Energy Physics Advisory Panel unanimously endorsed the recommendation of their Subpanel on New Facilities that this project should be given "the highest priority" in the US programme in order to achieve "completion in the first half of the 1990s" (Wojcicki [44, p.4]). If this target is met, the SSC would come into operation at about the same time as the earliest a proton machine could be completed in the LEP tunnel. Since the American facility would be capable of achieving energies up to twice those of the CERN machine, then, given the enormous cost of these machines, the arguments in favour of proceeding with the construction of such a CERN accelerator would be very substantially reduced. In short, the "real-estate" justification for LEP will lose a great deal of its force if construction of the SSC proceeds apace.

As was shown earlier in the right-hand column of table 1, the new proton synchrotrons at Fermilab and Serpukhov also provide scope for several possible future options, although not quite as many as with LEP. With the proton–antiproton colliders at CERN and Fermilab, however, the possibilities for future development are intrinsically more limited.[46] As before, these comparative advantages and disadvantages of the various new facilities have been summarized in table 5.

5. Overall assessment of the future prospects for CERN and its users

Having evaluated in detail the prospects for the various major new research facilities around the world, we are now in a position to come to some more general conclusions about the future for CERN and indeed for West European high-energy physics as a whole. In the short term, CERN's scientific prospects rest primarily upon the continued exploitation of the SPS and the proton–antiproton collider. To take the first of these, the SPS user-community has, since the late 1970s, enjoyed a range of advantages over the experimenters at Fermilab; these include a more reliable accelerator, higher-quality beams, technically more sophisticated detectors, and a higher level of financial support (including longer operating time – at Fermilab this was drastically curtailed during the early 1980s in order to limit the laboratory's electricity costs). However, now that the new Fermilab fixed-target machine has started to reach energies in excess of 500 GeV, the CERN SPS will gradually lose its position of world leadership for this type of accelerator, and its relative scientific output is likely to be affected accordingly.

---

[43] As was noted earlier, the construction of the first phase of LEP (LEP-50) will involve relatively conventional and well-established technology, with the result that the spin-off to accelerator physics is unlikely to be particularly significant.

[44] As Wilson [43, p.37] has observed, "this potentiality has not yet been publicly mentioned by the proponents of [LEP]", a situation that only began to change in 1984.

[45] The maximum energy would depend on the field-strengths achievable with superconducting magnets in the 1990s. 10 tesla magnets would be required to accelerate protons to an energy of 10 TeV in the LEP tunnel (cf. Wojcicki [44, p.26]).

[46] It should be noted, however, that the development of the CERN collider has paved the way for the low-energy antiproton ring (LEAR) facility. In addition, some thought has been given to the possibility of increasing the energy of the collider. However, this could only be achieved at the expense of substantially reducing the available luminosity, and, as we have seen, this is still rather low for certain types of important experiments.

As for the proton–antiproton collider, its advantages and disadvantages are well summarized in table 5. When CERN took the decision to embark on this project, the technical problems seemed formidable, and indeed there was an element of risk that the design luminosity would prove unattainable. Balancing this, however, was the fact that CERN had in the past two decades developed probably the greatest body of accelerator expertise among all the major world laboratories. As the project manager for one of the main US accelerator programmes remarked,

> On balance, I suppose our accelerator construction team ranks with CERN's first team, but they have second, third and fourth teams as well ... no-one can match that range of expertise. [Interview, 1981]

Given that CERN had acquired a record second to none in building and commissioning new accelerators, there were good grounds for optimism when the project began that the problems associated with the luminosity of the proton–antiproton collider would eventually be overcome, particularly when the likely lead over the Fermilab collider was stretched to some five years. While the CERN collider's relatively low event-rate might in principle be expected to restrict the range of physics that it can carry out, the big jump in energy it represents, and its comparatively large lead over the rival collider at Fermilab, should ensure that a fairly wide range of experiments is completed during the period up to 1986. Moreover, the CERN collider has proved relatively cheap in terms of the investment now required for a typical major new experimental facility, even if it has been built at the expense of small reductions in the reliability of the SPS and the time available for the fixed-target programme. This combination of relatively large scientific potential and low cost would suggest that, even though a high degree of technological uncertainty was involved, CERN chose wisely in 1978 when embarking upon this project – a conclusion shared by the great majority (93 percent) of the physicists whom we interviewed in 1981 and 1982.

In contrast, the comparative advantages and disadvantages of LEP are almost the exact opposite of these associated with the proton–antiproton collider, as is apparent from the pattern of strengths and weaknesses summarized in table 5. First, the technology required is relatively conservative and simple, apart from the tunnelling where civil engineering problems might possibly give rise to some delays. On the other hand, LEP has good potential for future development.

Second, the LEP project is extremely expensive compared with most new accelerators, and considerable criticism of the cost-effectiveness of the project was voiced by certain of the American physicists whom we interviewed (although some of those same physicists are now advocating the construction of the SSC costing many times more than LEP). The CERN Member States have agreed that they can find the funds, but only at the expense of substantial cutbacks in the laboratory's other research programmes – cutbacks of a size and significance not readily forseeable in 1977 when the European Committee for Future Accelerators first recommended that LEP should be CERN's next major machine. In particular, the Intersecting Storage Rings were closed at the end of 1983 – a major sacrifice given that the cancellation of ISABELLE meant this would have remained a unique accelerator until the mid-1990s and, moreover, one with relatively low operating costs.[47] Indeed, the evidence from interviews with physicists and from the bibliometric indicators presented in Paper II ([24] – see table 6 in particular) suggests that much worthwhile experimental work may still have remained to be done on the ISR.

Third, and last, whereas the proton–antiproton collider represents a significant increase in centre-of-mass energy, and one which will present its users with a position of monopoly in the energy region for perhaps five years, the first phase of LEP will achieve an energy no greater than that likely to be attained just over a year earlier by the

---

[47] There would have been a strong case for eventually developing the ISR as a heavy-ion collider, which would have produced another unique facility for a relatively small investment.

Stanford Linear Collider. True, LEP will have a certain advantage over the SLC in terms of luminosity, but it is far from clear how significant that advantage will prove. If the SLC reaches its design luminosity of 6×10^{30} cm^{-2} s^{-1}, it will produce over a million Z_0 events per year: in other words, the SLC should be able to carry out detailed studies (with reasonable statistics) of the production and decay-modes of this crucial particle (cf. [12, p.200]). This latter point has not always been recognized or admitted by the keenest supporters of LEP, as this CERN physicist's comments make clear:

> The local 'party line' at CERN is to doubt the SLC. They hope it won't work at all. But if it works within a factor of 10 or even 20 of the promised luminosity, and works in 1987, it will get a few thousand Z_0 decays. This will be a big psychological burden for LEP. Some people on the LEP project won't face up to this. The story then follows typically that at least LEP will be able to do precise measurements. Then people start talking about Phase II. But that costs a lot of money, and will have only a very low event rate. … So the usual CERN view is to say that the SLC won't work. Also one of the views on the LEP tunnel reconciling it is the 'real estate' view – the possibility of putting a proton accelerator and a proton–antiproton collider down there. This softens the view of the impact of the SLC as well. … I guess if I'm rational, I'd take measures to stop the crazy international competition. It has been terrible for SLAC losing the race with PETRA. If LEP is loser [in the competition with the SLC], this will be terrible for Europe. CERN *must* win the competition. But I'm pessimistic – the SLAC machine will probably just spoil LEP. [Interview, 1981]

As has been noted earlier, electron–positron colliders have a rather limited "physics menu" and therefore exhibit a tendency towards rapid obsolescence. Consequently, a lead of just one year for the SLC could prove extremely serious for CERN in that the research programme for LEP could consist very largely of repeating SLC experiments is a more detailed form. In other words, LEP might be restricted to the same sort of role as the CERN SPS performed in relation to the Fermilab 400 GeV accelerator. While such tasks as confirming discoveries, clearing up anomalies, and repeating experiments in order to obtain better statistics are clearly important, such work tends to have less impact than actually making a new discovery or carrying out the first measurement of a particular particle property. Thus, it is difficult for the outside observer to avoid the conclusion that the scientific potential of the first phase of LEP is likely to be limited by the SLC if the latter is completed on schedule and achieves a luminosity close to the design figure relatively quickly. In that eventuality, additional justification for the considerable investment made in the LEP project must be sought elsewhere – for example, in later phases of LEP and in possible proton facilities placed in the tunnel (but see the comments in section 4.5 above on the effect on the latter of the proposed SSC).

In view of these rather uncertain prospects for LEP, it is worth asking why enthusiasm for the project apparently remains so high. One possible explanation is that, just as the ISABELLE project could, to a large extent, be regarded as a historical relic of high-energy physics interests in the early 1970s [48] (before the technology was available to construct large proton–antiproton colliders), so LEP is a "child" of the mid-1970s when enthusiasm for circular electron–positron colliders was at its height. Then, a large storage-ring like LEP appeared to offer the best prospects for, if not discovering the Z_0 predicted by the newly triumphant Weinberg-Salam theory, at least investigating its properties in the "cleanest" and most systematic way. Since that time, the Stanford Linear Collider has been proposed – apparently capable of reaching the Z_0 region more quickly and considerably more cheaply – and the scientific promise of the proton–antiproton collider has been dramatically fulfilled. Nevertheless, much of the early enthusiasm for LEP still persists in public, although some physicists have begun to harbour private doubts (expressed to us during interviews) about whether the scientific case for LEP Phase I is still as strong as it was, say, in 1978. These doubts may at some stage surface in public if the pressure to secure completion of LEP by the planned date of 1988 continues to restrict severely the level of resources available for the laboratory's other experimental programmes.

Before we offer concluding remarks about the relevance of the present evaluation exercise for science-policy purposes, a few more general observations about the future for CERN are perhaps in order. Of special note is the fact that, apart from the Fermilab 1 TeV machine, all the main new Western facilities are colliders. This has profound implications for the extent and structure of experimental high-energy physics activity, since only a very limited number of experiments are possible at any one time on a collider. As a result, the present tendency for the size of collaborations to grow is almost certain to continue. In the case of LEP, where the intention of the CERN management is apparently to let anyone who wants to use the accelerator join one of the first four accredited experimental collaborations (cf. Robinson [37, p.42]), teams of perhaps 250 researchers now seem likely. In other words, a major fraction of Western Europe's 2000 high-energy physics researchers will be working on just four experiments, each with a lead time of six to seven years from the inception of the experiment to the production of scientific results. The same trend is likely to emerge in the United States, particularly if it is decided that the country can no longer afford to support three national high-energy physics laboratories.

---

[48] There is little evidence in table 5 to suggest that the United States was wrong to halt the construction of ISABELLE in 1983, despite the $200 million already spent.

The task of managing these vast collaborations will not be without its organizational and sociological problems: increasing bureaucracy; difficulties facing smaller universities in participating in what is an increasingly centralized activity; limitations on individual creativity and participation in the research process; difficulties in integrating postgraduate students into an experiment whose time-horizon stretches far beyond their training period; and so on. Indeed, it is the growing recognition of these problems that may result in increased pressures on CERN to continue supporting a full range of other experimental facilities, in particular to maintain a sizeable fixed-target programme, and in general to cater for those physicists unable or unwilling to work on LEP experiments. Nevertheless, the fact remains that the tasks of organizing the enormous experimental collaborations working on machines like LEP and providing them with the necessary support are likely to cause problems that will require careful management if they are not to threaten the scientific vitality of the large high-energy physics laboratories. In this respect, CERN probably has an advantage over the US centres since it has had more experience in dealing with such problems over the last 25 years. Hence, CERN is likely to suffer less from the effects of these problems on scientific performance. The advantage formerly possessed by the US laboratories over CERN by virtue of being able to react more flexibly and promptly to changing directions in the field (because of not having to cope with the time-consuming political demands associated with attempting to serve thirteen Member States) may disappear. With it may then go the success previously enjoyed by United States researchers in making most of the crucial discoveries in experimental high-energy physics during the 1960s and 1970s.

6. Some concluding remarks

As will already be clear, our overall assessment of the future for CERN is not a wholly optimistic one, particularly with regard to the scientific prospects for LEP. As such, it is likely to be subject to critical evaluation by LEP enthusiasts and others. No doubt, weaknesses in the analysis will be pointed out. This we fully expect, since any discussion of the future is inevitably clouded by uncertainties.

Yet besides analysing the future prospects for CERN, we would also hope that the paper will succeed in focusing attention on what many now believe are crucial deficiencies in the decision-making process in Big Science. A recent US report noted that

> there is some feeling that the high energy physics community does not have adequate opportunity to participate in this planning process which so affects its future. (Trilling [40, p.59])

However, perhaps of even greater concern is that those *outside* the specialty have even less say in decisions that can now involve the commitment of hundreds or even thousands of millions of dollars. It is perhaps somewhat surprising in an age where state expenditures have come under increasing scrutiny that a project like LEP with such long-term implications, and with such ramifications for other areas of science, was given the go-ahead with apparently so little real discussion in public or even within the scientific community (outside of high-energy physics). Moreover, the administrative device of classifying LEP as an "extension" of existing facilities, thereby curtailing debate about the merits of the project and indeed about the future of the CERN laboratory as a whole, is surely a somewhat debatable precedent from the point of view of publicly accountable science policy. Of course, it would be unfair to single out CERN for special attention in this respect; as was noted earlier, similar practices have been indulged in by US laboratories in protecting their institutional interests – for example, by funding new developments out of operating expenses until the resources committed reach such a level that virtually the only option remaining is completion of the project.

Such problems are, we would argue, intrinsic in those areas of Big Science characterized by large expensive central facilities that structure the interests of major sections of the scientific specialty concerned. Traditionally, the regulation of research activities and the formulation of scientific policies have depended on peer-review carried out in a relatively informal and qualitative manner. Yet under the conditions of "oligarchy" encountered in Big Science, it may become increasingly difficult to locate researchers capable of providing – and, more importantly, being seen to provide – the objective and disinterested judgements on which peer-review mechanisms depend.

It is in this context that we have explored the role of external assessments of the sort described above in helping to promote more open and rigorous discussion of policy options in Big Science. Obviously, the framework employed here for assessing the future prospects of major new research facilities is only one of the possibilities that might be considered for introducing some element of wider public and scientific participation into decision-making. Moreover, given that it is very much a first attempt to tackle this task, it may be subject to criticism. It is worth emphasizing, however, that the intention is *not* to replace the peer-review process – this must remain central in scientific decision-making – but to complement it, providing in a systematic and reproducible manner evidence that can inform decision-makers rather than determining their decisions. For example, if data similar to those appearing in table 5 above had not only been collected in 1978 but also made publicly available, it might have strengthened the case of those then arguing in favour of CERN pursuing the proton–antiproton option. Similarly, a comparison of such a table for 1978 with one, say, for 1981 would have revealed to high-energy physicists *and* the lay public alike that the likely scientific potential for LEP had been appreciably altered by the proposal of the Stanford Linear Collider in the intervening three years. This factor could then have been taken fully into account by politicians and science policy-makers in the CERN Member States (and not just by high-energy physicists) before any major expenditure on LEP had been incurred.

It should perhaps be stressed that the aim of the foregoing analysis is not so much to make CERN a stronger competitor in the international "race" with laboratories in the United States and elsewhere – the costs of the wastage involved in having "losers" as well as "winners" in such a capital-intensive activity as high-energy physics are now too great for even industrialized nations to afford – but rather to reduce the level of direct competition and duplication between major research laboratories and to encourage closer international coordination of research efforts. The latter is one of the concerns of the International Committee on Future Accelerators (ICFA), yet up till now this body has been largely thwarted in its efforts by the continued emphasis placed by physicists on "winning" the next international scientific "race". (This argument is generally most prominently deployed when government backing for a new project is being sought.) To take one example, in an ICFA workshop held in 1978, participants discussed new accelerator projects that were (apparently) beyond the means of individual regions (Eastern Europe, Western Europe, and North America) and which therefore required inter-regional collaboration. Most interest focused on a 20 TeV proton synchrotron, and various technical studies of such an accelerator were subsequently sponsored by ICFA. In 1983, however, US high-energy physicists began to urge their Government to build a proton machine of precisely this energy in order to ensure that their previous "dominant role" (Wojcicki [44, p.18]) in this field did not slip permanently overseas.[49] If plans for construction of this American machine do go ahead, then we can expect physicists in Western Europe (and presumably also in Eastern Europe) to counter in a few years time with plans for an even larger accelerator in order to ensure that the "lead" in high-energy physics is not once again lost to the United States.

---

[49] See also the discussion in Trilling [40, p.3] of the need for the United States to construct a substantial new facility over the next few years in order to maintain its "pre-eminence".

Just where this process of escalation (which has several features in common with the "arms race" between the super-powers) will end is by no means clear. Yet at some stage, the governments of the three "super-powers" in high-energy physics must surely call a halt on the grounds that the cost of the "next" accelerator (which will be measured in billions of dollars rather than millions) is too great for a single region to bear. High-energy physicists, who have for many years discussed vague plans for a "world accelerator" only for researchers in one region to persuade their government (or governments) to build such a machine and so steal a march on the other regions, will finally be forced to accept their own rhetoric that this particular field of fundamental intellectual endeavour is a truly international one in which national (or regional) considerations (such as attempting to wrest "the lead" from some other region) should play no part.[50] If this paper succeeds in stimulating discusion of this global "escalation" process and more generally of the problems associated with existing decision-making mechanisms in Big Science, it will have performed a useful function.

References

[1] W.J. Broad, Brookhaven Director Quits as ISABELLE Teeters, *Science* 213 (1981) 1089.

[2] Brookhaven Superconducting Magnet Operates at AGS, *CERN Courier* 13 (1973) 374–75.

[3] 1000 GeV Next Stop, *CERN Courier* 16 (1976) 12–13.

[4] People and Things, *CERN Courier* 16 (1976) 401.

[5] e-p Study Week, *CERN Courier* 17 (1977) 364.

[6] Projects Galore at Brookhaven, *CERN Courier* 18 (1978) 247–51.

[7] A Giant LEP for Mankind, *CERN Courier* 18 (1978) 431–435.

[8] The LEP Project, *CERN Courier* 20 (1980) 5–11.

[9] ECFA Meeting in May, *CERN Courier* 20 (1980) 191–193.

[10] KEK–TRISTAN Approval, *CERN Courier* 21 (1981) 103–104.

[11] Washington Conference, *CERN Courier* 21 (1981) 143–50.

[12] Stanford – Linear Collider Workshop, *CERN Courier* 21 (1981) 199–201.

[13] DESY – HERA Ahead, *CERN Courier* 21 (1981) 205–206.

[14] LEP Takes to the Hills, *CERN Courier* 21 (1981) 240–242.

[15] Stanford – Towards the New Collider, *CERN Courier* 22 (1982) 8–9.

[16] Klystrons Give 1 Megawatt, *CERN Courier* 22 (1982) 62–63.

[17] Novosibirsk – Preparing for VLEPP, *CERN Courier* 22 (1982) 417–418.

[18] KEK – TRISTAN Progress, *CERN Courier* 23 (1983) 3–5.

[19] Confidence at SLAC, *CERN Courier* 23 (1983) 81–82.

[20] DESY – HERA Clearer, *CERN Courier* 23 (1983) 90.

[21] Brookhaven – All the Way with CBA, *CERN Courier* 23 (1983) 127–128.

[22] Good News for LEP, *CERN Courier* 23 (1983) 228.

[23] J. Irvine and B.R. Martin, Assessing Basic Research: The Case of The Isaac Newton Telescope, *Social Studies of Science* 13 (1983) 48–86.

[24] J. Irvine and B.R. Martin, CERN: Past Performance and Future Prospects. II. The Scientific Performance of the CERN Accelerators, *Research Policy* 13 (1984) 247–284.

[25] J. Irvine and B.R. Martin, Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators, *Social Studies of Science* (forthcoming).

[26] J. Irvine and B.R. Martin, What Direction for Basic Scientific Research?, in: M. Gibbons, P. Gummett and B.M. Udgaonkar (eds.), *Science and Technology in the 1980s and Beyond* (Longman, London, 1984) 67–98.

[27] B.R. Martin and J. Irvine, An Evaluation of the Research Performance of Electron High-Energy Physics Accelerators, *Minerva* 19 (1981) 408–432.

[28] B.R. Martin and J. Irvine, Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy, *Research Policy* 12 (1983) 61–90.

[29] B.R. Martin and J. Irvine, CERN: Past Performance and Future Prospects. I. CERN's Position [illegible] World High-Energy-Physics, *Research Policy* 13 (1984) 183–210.

[30] W.D. Metz, Particle Physics: Many Results, Surprising Disclaimers, *Science* 178 (1972) 853.

[31] W.D. Metz, Two Superconducting Accelerators: Physics Spu[illegible]chnology, *Science* 200 (1978) 188–191.

[32] I. Miles and J. Irvine, Social Forecasting: Predicting the Future or Making History, in: J. Irvine, I. Miles and J. Evans (eds.), *Demystifying Social Statistics* (Pluto Press, London, 1979) 305–324.

[33] A.L. Robinson, International Competition Drives DESY, *Science* 212 (1981) 1488–1491.

[34] A.L. Robinson, CERN Sets Intermediate Vector Boson Hunt, *Science* 213 (1981) 191–194.

[35] A.L. Robinson, CERN Council Defers LEP Approval, *Science* 213 (1981) 528–531.

---

[50] Despite the public show of apparent unity among US high-energy physicists on the need to make the immediate initiation of the SSC project their top priority, some at least harbour doubts about the wisdom of the United States attempting to "go it alone" with such a vast undertaking: "My view is that high-energy physics should no longer be done on the basis of national facilities. We in the States should agree to help CERN put cryogenic magnets into the LEP tunnel [to make a 10 TeV proton synchrotron], and should collaborate technologically on this .... Then seven or eight years later Europe and the States should agree to do a big project in the US, for example, a big electron–positron collider. Proper *international* planning now needs to be done." [Interview, 1983].

[36] A.L. Robinson, DESY Looks to an International Future, *Science* 213 (1981) 530.

[37] A.L. Robinson, LEP Revolution Under Way at CERN, *Science* 217 (1982) 40–42.

[38] A.L. Robinson, CERN Gives Nod to Four LEP Detectors, *Science* 217 (1982) 722.

[39] A.L. Robinson, Fermilab Installing Superconducting Magnets, *Science* 217 (1982) 814–817.

[40] G. Trilling, *Report of the Subpanel on Long-Range Planning for the U.S. High Energy Physics Program of the High Energy Physics Advisory Panel* (Division of High Energy Physics, U.S. Department of Energy, Washington, D.C., DOE/ER-0128, 1982).

[41] L. Van Hove, The Research Activities of CERN (1976-1980) and the Future of the Laboratory, *CERN Annual Report 1980* (CERN, Geneva, 1980) 27–33.

[42] R. Walgate, On the Rocks, *Nature* 291 (1981) 275.

[43] R.R. Wilson, The Next Generation of Particle Accelerators, *Scientific American* 242 (January 1980) 26–41.

[44] S.J. Wojcicki, *Report of the 1983 Subpanel on New Facilities for the U.S. High Energy Physics Program of the High Energy Physics Advisory Panel* (Division of High Energy Physics, U.S. Department of Energy, Washington D.C., 1983).

[45] Fermilab Accelerator Conference, *CERN Courier* 23 (1983) 299–303.

[46] Panel Says: Go for a Multi-TeV Collider and Stop Isabelle, *Physics Today* 36 (September 1983) 17–20.

[47] Europe Uneasy at CERN Rival, *Nature* 309 (31 May 1984) 389–390.

• *ABSTRACT*

This paper presents the results of a study comparing the past scientific performance of high-energy physics accelerators in the Eastern bloc with that of their main Western counterparts. Output-evaluation indicators are used. After carefully examining the extent to which the output indicators used may be biased against science in the Eastern bloc, various conclusions are drawn about the relative contributions to science made by these accelerators. Where significant differences in performance are apparent, an attempt is made to identify the main factors responsible.

Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators

John Irvine and
Ben R. Martin[1]

Nations in both the East and West are increasingly being forced to consider how best to formulate appropriate and effective policies for science and technology.[2] One response has been the development of improved techniques for monitoring research outputs — whether they be contributions to scientific knowledge, or to technological innovation and economic development. Over the past five years, we have undertaken various studies of the scientific outputs from major research facilities in several Big Science specialties (radio astronomy, optical astronomy and electron high-energy physics). The overall aim of this work has been to develop and refine appropriate techniques for evaluating performance in basic research.[3] In this paper, we report the results of a comparative evaluation of the scientific outputs from accelerators in the Eastern bloc over the 1960s and 1970s. While we give particular attention to

Social Studies of Science (SAGE, London, Beverly Hills and New Delhi), Vol. 15 (1985), 293–341

the major accelerators at Dubna and Serpukhov, we also consider low-energy research facilities at four other laboratories, so that we cover a very large part of the Eastern bloc's effort in experimental high-energy physics. Our evaluation is based upon the method of 'converging partial indicators', described elsewhere;[4] it involves comparisons with similar facilities in Western Europe and the United States in terms of their outputs of research publications, the subsequent impact of those publications on the advance of scientific knowledge (as indicated by various citation statistics), and systematic peer-evaluation data obtained from interviews with scientists in Eastern and Western research institutes.

The paper begins by briefly reviewing recent Western work on the performance of Eastern-bloc science. Some major problems identified as restricting research in the Eastern bloc are discussed, along with the limitations of conventional sources of data on the subject — recollections of emigrés, reflections of visiting Western scientists, and international data-banks on publications and citations. These limitations are then taken into consideration when the differences in output between Eastern-bloc and Western accelerators are analyzed in detail. We argue that biases in the various indicators are, in fact, insufficient to account for the apparent differences in the output of similar research facilities in the East and West; this conclusion is supported by our peer-evaluation results. Finally, we attempt to identify some of the principal factors determining the scientific performance of Eastern-bloc accelerators, drawing on our interviews with a significant part of the world high-energy physics community.

The Growth of Interest in Eastern-Bloc Science

Two motives appear to underlie much previous work on Soviet and Eastern-bloc science. First is an obvious curiosity in the dynamics of a social system which, since 1917, has pursued a novel path of development, based in large part on policies attempting to make 'rational' use of all resources, including science and technology.[5] The Soviet revolution had profound ramifications for scientific research, and there was an upsurge of Western interest in the Soviet research and development system in the 1930s, following the congress on 'Science at the Crossroads' held in London in 1931.[6] Stalinist purges and the rise of such 'distorted' sciences as Lysenkoism provoked further interest. From this time on, and notwithstanding certain criticism, various groups of Western socialist scientists have drawn upon positive aspects of Eastern-bloc science to urge a fundamental reorientation of Western research towards a 'science for the people'.[7]

However, informed understanding of Eastern-bloc science really only began with, first, the exodus of emigré scientists with first-hand experience of working in research laboratories, and, second, the development (in the early 1960s) of exchange visits by scientists, of international collaborative projects,[8] and of conferences attended by researchers from East and West. This growth in personal contact allowed scientists to reflect upon the merits and weaknesses of each system, as well as opening up new opportunities for those studying comparative science policy[9] and the sociology of science.

A second motive is reflected in the emergence of systematic monitoring by the West of Eastern-bloc science, since this is now regarded as a key strategic element in industrial and military policy. Both at a public level (through organizations such as OECD), and at more covert levels, significant efforts have been devoted to studying the strengths and weaknesses of the main Eastern-bloc R&D sectors, and of the associated basic sciences. Such studies have not only considered the inputs to science and technology (funding levels, numbers of scientists and engineers,[10] and so on), but have also attempted to evaluate the outputs from R&D, particularly their potential value to military technology.[11] Political interest has, at times, clearly influenced the provision of government funds for academic studies of Eastern-bloc science.[12]

What We Know about Eastern-Bloc Basic Research Performance

Despite the recent proliferation of literature on Soviet and East European R&D, there is still remarkably little 'hard' quantitative data on the performance of its scientific system — certainly little that would be generally accepted in both East and West. Part of the problem is the reliance on personal recollections of emigré scientists which are, almost inevitably, strongly influenced by political values. The impression gained from emigré writings is one of chronic inefficiency, with the underlying cause located in an all-pervasive bureaucratic structure, the operation of which discourages creative and productive scientific research, especially among experimentalists.[13] The emphasis on long-term planning, in particular, is widely criticized on the grounds that it is difficult, if not impossible, to match the ever-changing and often unpredictable demands of experimental research with a relatively inflexible long-term plan based on prior identification of scientific problems. However, many other difficulties are also identified, most of them attributable to bureaucratic dysfunctions: these include those faced by creative young scientists in challenging conventional wisdom; the generally poor level of scientific management (in turn related to the over-centralized control of research); the stagnation of research institutes due to 'ageing' of their staff and a lack of mobility between laboratories; over-concern with secrecy; inadequate links between the Eastern-bloc scientific system and the rest of the world;[14] and, finally, poor links between research institutes and high-technology industry resulting in the stunted development of many areas of experimental science denied modern instrumentation and adequate computing facilities. In short, emigrés argue, most areas of basic research (with only a few exceptions) are uncompetitive with the West because the conditions essential for successful research are generally lacking.

From the point of view of comparative science-policy analysis, the opinions of visiting foreign scientists have one advantage over emigré writings: although visitors clearly have less knowledge about the detailed operation of the scientific system, they are more likely to hold a balanced view.[15] While visitors tend to confirm reports of obsolete experimental equipment and computing facilities, a poor scientific communication system, an over-developed bureaucracy, and weak links between scientific research and technology-based industry, they also recognize that in several areas of science (particularly the theoretical branches) the performance of Eastern-bloc researchers has been amongst the best in the world, at least partly because young scientists are given a thorough grounding in theory.

Prominent among the initiatives aimed at promoting collaborative research has been the US-USSR Inter-Academy Exchange Program, which has been reviewed (together with other aspects of American-Soviet scientific collaboration) by the US National Academy of Sciences.[16] American scientists who had visited the USSR were asked in a postal questionnaire to assess their experiences, and to comment upon the quality of Soviet scientists and the strength of their respective disciplines.[17] This approach yielded useful qualitative peer-evaluation data on a field-by-field basis; for example, Soviet theoretical physicists and mathematicians were judged to be among the very best in the world, while particle physics was generally regarded as lagging 'between five and ten years behind that in the United States and Western Europe'.[18] Important quantitative information was also provided on American scientists' perceptions of the overall performance of Soviet science; just under 80 percent regarded the Soviet Union as less advanced scientifically than the United States. However, little attempt has been made to relate such opinion surveys to more systematic evaluations. Moreover, the extent to which the views of American scientists accord with those of their Soviet colleagues remains a matter for speculation, since the Soviet Union decided it would be inappropriate for a parallel survey of scientific opinion to be undertaken by the Academy of Sciences in the USSR. Nevertheless, as L. R. Graham (rapporteur to the NAS review) has pointed out, 'this is the first attempt made by qualified American scientists to evaluate Soviet science . . . systematically.'[19] As such, it represents a valuable addition to the literature on Soviet science.

A further source of data is the series of macro-level statistics on publications and citations produced by the Institute for Scientific Information (ISI), as a by-product of their *Science Citation Index* (*SCI*). Attempts to use such data to compare research outputs have been critically received by many Western scientists, who tend to prefer traditional methods for assessing research performance based on expert peer-review. Yet the analysis of publication and citation data has become widely used as a method of comparing national scientific outputs, particularly since the appearance in *Science Indicators 1972* of field-by-field counts of the scientific papers published by the major industrialized countries, and figures on the frequency with which they are cited.[20] According to those who produced these data,

> . . . publication counts were used as indicators of national scientific activity, while the citation counts were used as indicators of national scientific 'quality' or 'significance'.[21]

On these indicators, the USSR compares relatively poorly with most other major countries, and in particular the US. Such findings are now integrated into the conventional wisdom of Western science policy, as the following statement reveals:

> By any measure — whether Nobel prizes, frequency of citation by fellow specialists, origin of major breakthroughs, or simply quantity of publications — US scientists lead their Soviet colleagues in most disciplines, and in many there is simply no competition.[22]

Uncritical acceptance of these figures as indicators of national scientific performance is, however, extremely dangerous, since the ISI data-base has intrinsic methodological limitations. This is especially true for comparisons between East and West: the coverage of journals in the *SCI* is very unevenly distributed across countries, with English-language nations faring particularly well. In physics, for example, while 160 Soviet and 191 US physics journals were held in 1973 by the extremely comprehensive British Library Lending Division,[23] only 14 of the Soviet journals (less than 10 percent) compared with 40 for the US (over 20 percent) were scanned by ISI — a discrepancy also evident in abstract counts.[24] Such bias is likely to lead to underestimates of Soviet publication output. If physics is typical, then Soviet journals are under-represented by a factor of just over two. A recent study of the adequacy of the ISI data as an indicator of relative international scientific activity was forced to conclude that in 'the case of the USSR, the coverage is so deficient that international activity indicators based on this Soviet coverage would be seriously affected'.[25] Bias is also inevitably present in the citation data. If less than 10 percent of all Soviet (but over 20% of US) physics journals are scanned by ISI, then Soviet physics publications are 'losing' a far larger number of their citations (in the unscanned journals) than are American ones. ISI-based data purporting to reveal the low 'quality' of Eastern-bloc science should thus be regarded with the same scepticism as data on relative numbers of publications.

The overall conclusion is that the evidence available on the performance of Eastern-bloc science is not at present sufficiently reliable to support the critical views often heard in the West. Publication and citation data apparently corroborate the views of emigré scientists, the impressions gained by Western visitors, and the general findings of Western surveys of R&D output. Yet this does not necessarily mean that the collective picture is a faithful one, since these sources all have major biases and technical limitations. Is it possible to conduct more systematic studies of scientific

AVAILABLE

performance that are less susceptible to 'bias', and which would, therefore, be as acceptable to the scientific community and science-policy analysts in the Eastern-bloc countries as they are in the West?

An Alternative Framework for the Evaluation of Research Performance

The results of work at the Science Policy Research Unit over the last seven years suggest that such evaluation techniques can indeed be developed. In what follows, we describe an approach, based on a number of 'partial indicators' of scientific progress, for making comparisons of major experimental facilities (high-energy physics accelerators) and of the scientists who use them.

High-energy physics exemplifies 'curiosity-oriented science' — that is, research funded primarily because of the contributions it is likely to make to the advance of knowledge. It also leads to substantial educational benefits (in the form of highly trained scientific personnel), to various types of technological 'spin-off',[26] and even to broader political benefits such as increased national prestige and improved international cooperation. Yet these secondary reasons cannot explain why, over the last decade, nations have spent around $1,000m a year on the subject.[27] Our assessment therefore focuses on the contributions to scientific knowledge ('scientific' contributions) associated with different high-energy physics accelerators.

In previous work, we have developed a 'method of converging partial indicators' for comparing scientific contributions. The method employs various partial indicators of the scientific progress made by users of different research facilities: publication counts, citation statistics (total citations, citations per paper, and numbers of highly cited papers or 'discoveries'), and extensive peer-evaluation rankings of the relative outputs of the facilities. These indicators can be applied to matched groups of researchers using similar research facilities and publishing in essentially the same body of journals. In each of the three Big Science specialties in which we have used the method, a certain convergence between the results based on each partial indicator has been obtained; reasonably unambiguous conclusions can then be drawn about the relative contributions to scientific progress from each experimental facility.[28] The main elements of the method are summarized in Table 1.

TABLE 1

Main Problems with the Various Partial Indicators of Scientific Progress and Details of how their Effects may be Minimized Using the Method of Converging Partial Indicators

Partial Indicator Based on	Problem	How Effects May be Minimized	
A *Publication Counts*	1 Each publication does not make an equal contribution to scientific knowledge	Use citations to indicate average impact of a research facility's publications, and to identify very highly cited papers	Use only indicators that yield convergent results
	2. Variation of publication rates with specialty or institutional context	Choose matched research facilities producing similar types of papers within a single specialty	
B *Citation Analysis*	1 Technical limitations with *SCI*:		
	(a) first author only listed;	Not a problem when dealing with a research facility	
	(b) variations in names; (c) authors with identical names; (d) clerical errors;	Check manually	
	(e) incomplete coverage of journals	Not a serious problem for Big Science	
	(2) Variation of citation rate during lifetime of paper — unrecognized advances on the one hand, and integration of basic ideas on the other 3 Critical citations 4 'Halo effect' citations	Not a problem if citations are regarded as an indicator of impact, rather than quality or importance	
	5. Variations of citation rate with type of paper and specialty	Choose matched research facilities producing similar types of papers within a specialty	
	6. Self-citation and 'in-house' citation (SC and IHC)	Check empirically and adjust results if the incidence of SC or IHC varies between groups	
C *Peer Evaluation*	1. Perceived implication of results for own central facility and competitors may affect evaluation 2. Individuals evaluate scientific contributions in relation to their own (very different) cognitive and social locations 3 Conformist assessments (e.g. 'halo effect') accentuated by lack of knowledge on contributions of different research facilities	1 Use large representative sample 2. Use verbal rather than written survey in order to press evaluator if a divergence between expressed opinions and actual views is suspected 3 Assure evaluators of confidentiality 4. Check for systematic variations between different groups of evaluators	

Our approach in this study was to identify the main Eastern-bloc experimental high-energy facilities, and then to compare their scientific outputs with those from the nearest equivalent facilities in the West. The data-base for the publication and citation indicators has been carefully examined to assess the extent of any bias against Eastern-bloc science, and hence to overcome, as far as possible, the main problems with previous use of such data. These bibliometric data are complemented by a survey of scientific opinion on the relative performance of Eastern-bloc accelerators, involving interviews with nearly 200 high-energy physicists in both East and West; again, the results were carefully analyzed for any systematic bias. Although there are definitely biases in all the output indicators, one can estimate their approximate magnitudes, and so allow for their effects. The indicators, where they converge, can then be used to draw what we believe are reasonably accurate conclusions about the relative performance of Soviet and East European experimental high-energy physics. There appears to be no reason in principle why such an evaluation could not be applied to other areas of Eastern-bloc science.

Eastern-Bloc Accelerators and their Closest Competitors

High-energy physics has been comparatively well supported in both East and West. In Western nations, it has at times accounted for 40 percent or more of total state expenditures on basic natural science.[29] Initially, during the 1950s and early 1960s, this favoured position seems to have stemmed from an implicit belief that significant contributions would be made to nuclear-energy research.[30] When this belief later became untenable, high-energy physicists began to emphasize the 'fundamental' nature of the field, and its potential contributions to other sciences.[31] In the Eastern bloc, such considerations have gone hand in hand with matters of national prestige. (Great effort, for example, was put into ensuring that the Serpukhov facility, then the highest-energy accelerator in the world, was ready in 1967 to celebrate the fiftieth anniversary of the Russian revolution.) As a result, according to the senior American high-energy physicist, R. R. Wilson, 'the scope of the accelerator laboratories in the USSR is comparable to that of the American and European programs'.[32]

Table 2 lists the main Eastern-bloc accelerators operating from

TABLE 2

Eastern-Bloc Accelerators and Their Closest Competitors, 1960–82

Accelerator	Date of First Operation	Beam Energy (GeV)	Approximate Cost (US $m)[a]
PROTON >25GeV			
Serpukhov (USSR)	1967	76	~100
CERN PS (W. Europe)	1959	28	45
Brookhaven AGS (USA)	1960	33	31
CERN ISR (W. Europe)	1971	31	75
Fermilab (USA)	1972	500	250
CERN SPS (W. Europe)	1976	500	250
PROTON <25GeV but >5Gev			
Dubna (E. Europe)	1957	10	~30?[c]
Moscow (USSR)	1961	7[b]	~20?[c]
Berkeley (USA)	1954	6	34
Argonne ZGS (USA)	1963	13	50
Rutherford (UK)	1963	7	28
ELECTRON >5GeV			
Yerevan ARUS (USSR)	1967	6	~10?[c, d]
Cambridge CEA (USA)	1962	6	12
Hamburg DESY (FRG)	1964	7	19
SLAC (Stanford) (USA)	1966	22	114
Daresbury NINA (UK)	1966	5	12
Cornell (USA)	1967	12	12
LINEAR e^- <5GeV			
Kharkov (USSR)	1964	2	~10?[c]
Stanford Mk III (USA)	1952	1.2	7
Orsay (France)	1959	2.3	14
e^+e^- COLLIDERS			
Novosibirsk (a) VEPP 1	?	0.16	?
(USSR) (b) VEPP 2	1965	0.5	~2?[c]
(c) VEPP 2M	1974	0.7	?
(d) VEPP 3	1970	3	?
(e) VEPP 4	1979	7	~20?[c]
Orsay (a) ACO	1965	0.5	2
(France) (b) DCI	1976	1.8	13
Frascati[e] ADONE (Italy)	1969	1.5	9
Stanford[e] SPEAR (USA)	1972	4	~24
Hamburg DORIS (FRG)	1973	5	45

[a] Actual prices (unadjusted for inflation) based on official exchange rates.

[b] Subsequently raised to 10 GeV.

[c] Estimates only, based on the costs of similar Western machines.

[d] A smaller Soviet electron synchrotron at Tomsk cost two million roubles. Since the capital cost of an accelerator increases approximately linearly with energy, one would expect the Yerevan electron synchrotron with five times the energy of the Tomsk accelerator to have cost approximately five times as much — i.e. ten million dollars, if one assumes an exchange rate of 1 rouble = 1 US dollar.

[e] This laboratory did operate a smaller collider in the early 1960s, but it was used more as a test-bed for accelerator physics than as an experimental high-energy physics research facility.

1960 to 1982, together with their nearest equivalent Western competitors. The list includes virtually all the world's largest accelerator facilities.[33] All the Eastern-bloc facilities are in the Soviet Union, while the main Western accelerator centres are in the US and Western Europe. Since 1967, the largest Eastern-bloc facility has been the Soviet 76 GeV (billion or giga-electron volts) proton synchrotron operated by the Institute for High-Energy Physics (IHEP) at Serpukhov. From 1967 to 1972, this was the most powerful accelerator in the world, just as the 10 GeV synchrophasotron at the international Joint Institute for Nuclear Research (JINR) at Dubna had been for a brief period towards the end of the 1950s.[34] Another major proton accelerator is the 7 GeV (subsequently upgraded to 10 GeV) machine operated by the Institute for Theoretical and Experimental Physics (ITEP), Moscow. As for the electron machines, there are the 6 GeV synchrotron at Yerevan in Armenia, the 2 GeV linear accelerator at Kharkov in the Ukraine, and various electron-positron colliders at Novosibirsk in Siberia. Table 2 classifies the various accelerators according to energy-range and particles accelerated, this being the most obvious way of identifying the main sets of competing facilities. Thus, the Serpukhov accelerator is most directly comparable with the slightly older and lower-energy synchrotrons at CERN (the Proton Synchrotron or CERN PS) and Brookhaven National Laboratory (the Alternating Gradient Synchrotron or AGS) on the US East Coast, as well as the CERN Intersecting Storage Rings (ISR) and the 'Super' Proton Synchrotrons at CERN (the CERN SPS) and Fermilab in the US Mid-West. Similarly, the Dubna and Moscow proton accelerators can be compared with the machines at Argonne National Laboratory near Chicago[35] and Rutherford Laboratory in Britain, with the slightly older Bevatron at Berkeley in California, and to a certain extent with the somewhat larger but contemporary CERN PS and Brookhaven AGS. Among electron machines, the Yerevan synchrotron had several virtually identical rivals at Cambridge in Massachusetts, at the German Electron Synchrotron Laboratory (DESY) near Hamburg, and at Daresbury Laboratory in Britain.[36] The higher-energy synchrotron at Cornell University and the linear accelerator at the Stanford Linear Accelerator Center (SLAC) in California also provide useful comparisons. The Kharkov electron linear accelerator is very similar to machines operated for many years by Stanford University (the Mark III), and by the Orsay Laboratory in France. Finally, the

Novosibirsk electron-positron colliders can be compared with those operated by Orsay, Frascati Laboratory (in Italy), SLAC, and DESY.

How great have been the scientific contributions from these various Eastern-bloc accelerators, especially from the principal machine at Serpukhov, compared with those from their Western counterparts? A recent official US Congress Report concluded (on the basis of rather limited evidence) that 'to date, few significant physics discoveries have been attributed to research conducted on Soviet accelerators.'[17] How accurate is this assessment?

Accelerator Outputs

Scientific Publications

Publication counts are one of the most frequently used, but also most problematic, indicators of scientific output. Their use as an output indicator rests on the assumption that scientists (in basic research, at least) prefer to present the end-result of most research in scholarly publications. Not only is this thought to be an effective way to transmit information to the scientific community, but it also constitutes an important means of securing recognition in that community; furthermore, it acts as a convenient accounting device for funding agencies in ensuring that scarce resources have been effectively used.

Because of the differing institutional, social and political pressures to publish in different countries, publication counts need to be undertaken with considerable care (both technical and conceptual).[18] It must be recognized, for example, that publications do not all have an identical impact on the advance of scientific knowledge, and that the same results are sometimes republished in different forms (often first in conference proceedings, and later in a journal). Moreover, in attempting to compare the scientific outputs of different research facilities, one must ensure that only 'like' is compared with 'like'; for example, the apparently greater publication output of one high-energy physics laboratory compared with a rival operating similar research facilities may be because the former has a large theory group attached (theorists tend to publish rather more frequently than experimentalists).[19]

In comparing publication outputs, we have tried to ensure that our methods are as unbiased as possible. Only high-energy physics papers reporting new (that is, previously unpublished) experimental data have been included. Preprints and conference papers were excluded since revised versions are generally published later. And, since we are only comparing experimental outputs (which provide the rationale for constructing costly accelerators), we have excluded all theoretical, instrumental and review papers. We also attempted to ensure full coverage of all the main Eastern-bloc and Western journals. This was done by asking physicists to identify the journals used to publish the bulk of experimental high-energy physics results; analysis of the results showed that most papers are published in eleven major international journals (listed in note (b) to Table 3). Four of the journals are Soviet, five are West European, and two are American. We scanned all these journals from 1961 to 1982, classifying experimental high-energy physics papers in terms of the accelerator (or accelerators — see note (c) to Table 3) used.[40] The resulting publication list was then cross-checked against various data compilations and lists of papers provided by research laboratories. This search yielded about five percent more items: some of these involved borderline decisions on whether the results reported were 'experimental' or 'high-energy physics' (if there was any doubt, they were included); the remainder were in journals other than the eleven scanned. The final totals for each accelerator are shown in Table 3.

Before examining the figures, we must look critically for possible sources of bias. One may arise from journals that have not been scanned — in particular, national physics journals and those of individual Soviet Republics. Whereas such journals probably account for 10 percent or less of all experimental high-energy physics publications in the West, the corresponding figure in the East is likely to be rather higher. This is especially true for smaller accelerators such as those at Kharkov and Yerevan (the users of which sometimes publish in the *Ukraine Physics Journal* and the *Proceedings of the Armenian Academy of Sciences* respectively), and for Dubna, where visitors from Eastern Europe publish at least some papers in domestic physics journals. The effect is probably greatest for Kharkov, where we estimate that up to 25 percent of papers are published in unscanned journals; but for the Serpukhov, Moscow and Novosibirsk facilities, the figure appears to be only 10 to 15 percent,[41] much closer to the value for Western accelerators. Although we have made some attempt to include such publications by using various other publication lists (see note (b) to Table 3), our

BEST COPY AVAILABLE

TABLE 3

Numbers of Experimental High-Energy Physics Papers[a] Published in International Journals[b] during the Preceding Two Years

Accelerator[c]	1962	1964	1966	1968	1970	1972	1974	1976	1978	1980	1982
PROTON >25GeV											
Serpukhov	–	–	–	–	29	63	109	123	136	129	110
CERN PS	37	157	211	215	236	254	229	278	246	108	67
Brookhaven AGS	17	58	127	166	178	150	153	93	60	51	31
CERN ISR	–	–	–	–	–	21	38	58	51	51	62
Fermilab	–	–	–	–	–	5	107	176	182	179	116
CERN SPS	–	–	–	–	–	–	–	–	26	78	130
PROTON <25GeV											
Dubna	69	64	42	33	34	22	20	29	36	24	19
Moscow	7	15	15	27	24	24	43	36	24	24	15
Berkeley	143	122	82	95	83	60	34	13	10	5	3
Argonne ZGS	–	–	13	55	88	97	102	68	57	29	21
Rutherford	–	–	17	21	23	26	32	20	20	22	2
ELECTRON >5GeV											
Yerevan ARUS	–	–	–	–	2	10	7	7	9	11	10
Cambridge CEA	–	11	16	35	22	10	8	–	–	–	–
Hamburg DESY	–	–	11	39	24	37	33	22	17	4	1
SLAC	–	–	–	21	54	82	74	79	68	33	15
Daresbury NINA	–	–	–	–	9	18	13	10	10	12	5
Cornell	–	–	–	–	14	17	14	18	18	8	3
LINEAR e^- <5GeV											
Kharkov	–	–	–	6	20	31	23	29	18	22	14
Stanford Mk III	14	19	22	18	12	6	2	1	–	–	–
Orsay	2	11	11	11	5	4	3	–	–	–	–
e^+e^- COLLIDERS											
Novosibirsk	–	–	–	3	1	9	1	0	6	4	10
Orsay ACO + DCI	–	–	–	1	7	4	4	4	2	6	9
Frascati ADONE	–	–	–	–	5	16	26	17	12	10	5
Stanford SPEAR	–	–	–	–	–	–	4	30	36	21	25
Hamburg DORIS	–	–	–	–	–	–	1	10	30	16	10

a. 'High-energy physics' is defined here as physics carried out with accelerators able to produce primary particles at an energy higher than 1 GeV. 'Experimental' high-energy physics papers are those which contain new (i.e. previously unpublished) experimental data. We have excluded theoretical papers, reviews or compilations of data, preprints, book articles, conference proceedings and theses. Also excluded are papers on instrumentation, nuclear level structure, and studies of cosmic rays.

b. This publication list was derived by scanning the following journals: *Soviet Physics-JETP*; *JETP Letters*; *Soviet Journal of Nuclear Physics*; *Soviet Physics-Doklady*; *Nuclear Physics B* (and before that, *Nuclear Physics*); *Physics Letters B* (and before that, *Physics Letters*); *Nuovo Cimento*; *Lettere al Nuovo Cimento*; *Physical Review D* (and before that *Physical Review*); *Physical Review Letters*; and *Zeitschrift für Physik C*. Additional information came from Annual Reports and publication lists provided by the various laboratories, and data compilations such as Particle Data Group, *An Indexed Compilation of Experimental High-Energy Physics Literature* (Berkeley, Calif.: Lawrence Berkeley Laboratory, LBL-90, 1978). As a result, the final publications list contains a small number of papers published in a variety of other journals.

c. All the papers were scanned to establish which accelerator was used to obtain the experimental results. In a small number of cases (2.4%), more than one accelerator was used. For such cases, each accelerator used was credited with that paper.

coverage of them is still rather incomplete. When the overall coverage of papers from Western accelerators (90 to 95 percent) is compared with that for papers from Eastern-bloc facilities (between 75 and 90 percent), it can be seen that a bias of some 10 to 15 percent remains against the latter (and rather more in the case of Kharkov).

A second potential source of bias lies in differing publication practices. Pressures to publish may be greater in the West (because of the use of publications to determine promotion and project funding), and this may lead researchers to break up their experimental data into several short articles when a single substantial paper would suffice, or to publish results prematurely to achieve 'priority'. In the East, priority is achieved by submitting work to state committees charged with scrutinizing research results before publication can take place. The long process of publication, beginning with formal defence of work within the laboratory, and the need to obtain the signatures of senior administrators, may thus not only reduce the relative number of publications but improve their overall quality. However, such effects should be revealed in the citation data, since one would expect longer, more substantial, and higher-quality papers to earn more citations. (This question is considered later in note 54.)

A third source of bias is that it has become customary for certain Eastern-bloc researchers to publish their results twice — once in Russian in a Soviet journal, and once in English in a West European one. This particularly affects Serpukhov, and, according to physicists interviewed in Eastern Europe, tends to arise most in East-West collaborations. We estimate that this introduces a bias *in favour* of Serpukhov of 10 percent or so,[42] which almost cancels out the first source of bias discussed above.

Finally, our figures are in general biased against those facilities that are used for types of research other than high-energy (particle) physics (such as nuclear physics or synchrotron-radiation work.) This applies particularly to the three electron linear accelerators, and in recent years has also become the case with the proton accelerators at Dubna, Moscow and Berkeley. Thus, in evaluating the *high-energy physics* outputs of these accelerators, we are considering only part of the research work they undertake. (The distribution of activity at such facilities around the world is probably much the same, so this source of bias may not be particularly important.)

Given these reservations, what do the figures in Table 3 tell us?

Although the annual output of papers from Serpukhov has been just over twice as great as that from the CERN Intersecting Storage Rings,[43] it was appreciably less than from not only the larger Fermilab accelerator, but also the lower-energy machines at CERN and (initially) at Brookhaven (until its productivity plummeted in the mid-1970s as American experimentalists migrated to the more powerful facility at Fermilab). Similarly, the outputs from the Dubna, and particularly the Moscow, accelerators were in general considerably smaller than those from the nearest equivalent American machines at Berkeley and Argonne (and only a fraction of those from the somewhat higher-energy CERN PS and Brookhaven AGS accelerators), although they were roughly similar to that from the British accelerator at Rutherford Laboratory. Such an imbalance also appears for the world's main electron accelerators. The synchrotron at Yerevan produced relatively few papers, less even than the Daresbury machine. Indeed, of all the Eastern-bloc accelerators, only that at Kharkov appears to have produced a comparatively high number of research papers (though this may be less a reflection of its relative efficiency than the fact that, by the early 1970s, the interest of high-energy physicists in the West had moved on from the rival facilities at Stanford University and Orsay to other accelerators).

Even allowing for the possible bias of up to 15 percent discussed earlier, it must be concluded that, overall, the publication output from the Eastern-bloc accelerators is low in world terms — typically a factor of two below that of equivalent Western facilities. In particular, neither Dubna nor, to a lesser extent, Serpukhov, seems to have capitalized on their temporary positions as the highest-energy machines in the world in the same way that, in their respective periods, the CERN PS and Brookhaven AGS, and later the Fermilab accelerator, appear to have done. However, publication counts on their own are only a very limited indicator of scientific contributions. We need to know more about the relative impact on the advance of scientific knowledge of these papers. It may well be that, because the career structures of Eastern-bloc scientists are less subject to the 'publish or perish' syndrome, they have tended to publish only more substantial papers. If the contribution to scientific progress made by Eastern-bloc papers has, on average, indeed been higher than that by Western papers, then this might offset the generally much lower publication output of Eastern-bloc accelerators. Is there any evidence for such a countervailing effect?

Overall Impact of Scientific Publications

In an attempt to provide a means of 'weighting' publications according to their relative contribution to the advance of knowledge, science-policy analysts have in recent years turned to citation data. However, as we have argued elsewhere,[44] citation figures provide not so much an indicator of the 'quality' or 'importance' of papers, as of their 'impact' on the advance of knowledge. A high-quality paper in a stagnating field may contribute little to the advance of knowledge, and hence receive less citations than a paper of similar quality in a more active field. A paper's 'importance' is the influence it would achieve were scientific communication completely free from institutional, social and political constraints. Thus, a potentially very influential paper may go unnoticed and uncited if it is written in an obscure journal by a not-very-prominent author, or in a non-English language. A publication's 'impact', in contrast, describes its *actual* influence on the advance of knowledge, and it is this for which citations provide a (partial) indicator.[45]

The citation records of each of the experimental publications in our list were obtained by manual scanning of the *SCI* for the years 1961–82. Unlike computer scanning, this method enables most of the technical problems associated with the *SCI* (misspelt names, incorrect references, and so on — see Table 1) to be easily overcome. Since the *SCI* covers the eleven main international high-energy physics journals, as well as most of the subsidiary journals occasionally used by Western high-energy physicists, the citation data for accelerators in the West should be between 90 and 95 percent complete (like the publication totals). In contrast, the citation data for Eastern-bloc accelerators are almost certainly much less complete.

There are three main sources of likely bias. First, there are the papers in our list that 'lose' citations from articles in journals not scanned by ISI. We have seen that up to 20 percent or so of Eastern-bloc papers are published, not in the eleven major international journals, but in the national physics journals of either East European states (which are generally scanned by ISI), or of individual Soviet Republics (which are not). Citations from articles in these latter journals are 'lost'. On the basis of the estimate that 10 to 20 percent of high-energy physics papers in the Eastern bloc are published in unscanned journals, we can expect a bias of similar magnitude in the

AVAILABLE

citation counts,[46] although it is likely to be under 10 percent for the more international accelerators (Serpukhov and Dubna) since these earn a larger fraction of their citations from Western journals, nearly all of which (for this field at least) are scanned by ISI.

A second bias involves those experimental publications omitted from our list altogether; these 'lose' all their citations. If such publications were on average cited with the same frequency as publications in the list, the resulting citation bias against Eastern-bloc accelerators would be of the same magnitude as that in the publication data[47] — that is, 10 to 15 percent, and slightly more than 20 percent in the case of Kharkov and Yerevan. However, it is clear from our data on papers published in the physics journals of Soviet Republics that these are cited far less often than papers in the international journals (which have a larger audience and therefore reach more potential citers).[48] Hence, this particular bias is somewhat smaller, perhaps 10 percent or less.[49]

A third possible source of bias, which is rather harder to estimate, is that papers published in the main Soviet journals tend to contain less references on average than those in equivalent Western journals. Full-length Soviet papers typically contain around three-quarters as many references as Western articles, and 'letters' only half as many.[50] There are three possible explanations: either Eastern-bloc physicists draw less heavily on previous results because such work has less impact on them; or they cite fewer references because their own literature is not so extensive as that of the West, and therefore contains fewer potentially citable articles; or they are more discriminating in their use of references. If the last is true, then each Eastern-bloc citation implies a greater degree of indebtedness than a Western citation. This would suggest that, in comparing the relative scientific impact of research between East and West, one should 'weight' Eastern-bloc citations more heavily. Assuming an average figure (for both full-length articles and 'letters') for the East/West 'weight' ratio of about two-thirds implies that 'unweighted' citation data will be biased by 33 percent against the Eastern bloc. However, Eastern-bloc research earns a substantial fraction of its citations from articles in Western journals, thereby reducing the effect to nearer 20 percent. Even this is probably an over-estimate of the residual bias, since the first explanation of the lower number of references in Eastern-bloc papers (previous work having less impact) is at least partly valid. Moreover, the bias should be considerably smaller for work of international interest (such as that at Serpukhov), since this will be referred to more equally by Western and Eastern authors.

Taking all three sources of bias together,[51] it appears that an overall citation bias against Eastern-bloc accelerators in the region of 30 to 50 percent might be expected; the larger machines probably fall in the lower part of this range, and the smaller ones in the upper part. This means, for example, that the 'crude' citation figures for the Kharkov accelerator should be doubled, while those for Serpukhov or Dubna should be increased by just under a half. Such biases should be borne in mind when examining Table 4.

Table 4 presents the numbers of citations achieved by each accelerator over four-year periods from 1961 onwards as listed in even-year editions of the *SCI*. The figures suggest that publications from Serpukhov had the greatest impact between 1974 and 1976, these citations being to experiments in the early 1970s when the accelerator was briefly the world's highest-energy machine. Overall, however, the impact of Serpukhov publications over the accelerator's first twelve years appears to have been very much less (even allowing for a possible 30 percent bias) than papers from the lower-energy CERN PS and Brookhaven AGS accelerators over their first twelve years, and even smaller than the impact of papers from the Fermilab accelerator. Similarly, the number of citations earned by publications from the Dubna and Moscow accelerators is approximately an order of magnitude less than that for the CERN PS and the Brookhaven AGS, even after allowing for bias. These two Eastern-bloc machines have also made considerably less impact than similar accelerators at Berkeley and Argonne, and probably slightly less even than the Rutherford Nimrod accelerator. Similarly, the impact of the Yerevan accelerator has been an order of magnitude less than all the other large electron accelerators (except for Daresbury). The various Novosibirsk machines have done little better, with an impact an order of magnitude less than those at Stanford and Hamburg. The gap between the Kharkov accelerator and its rival is somewhat smaller, but this may be because in the West this particular energy-region was fairly thoroughly explored during the 1960s, and because by the early 1970s the interests of experimenters had switched elsewhere.

Overall, then, we conclude that, although the Serpukhov accelerator achieved a relatively high impact during the first half of the 1970s, the evidence points to this having been exceptional among Eastern-bloc accelerators. Even allowing for the various sources of

TABLE 4
Numbers of Citations to Experimental Journal Articles Published During the Preceding Four Years

	1964	1966	1968	1970	1972	1974	1976	1978	1980	1982
PROTON >25GeV										
Serpukhov	–	–	–	139	273	482	494	318	311	301
CERN PS	375	1280	1327	1195	1153	1322	1135	1174	774	250
Brookhaven AGS	597	1265	1350	1202	959	757	779	419	174	116
CERN ISR	–	–	–	–	271	825	706	590	443	528
Fermilab	–	–	–	–	19	778	1903	2622	1318	804
CERN SPS	–	–	–	–	–	–	–	317	523	801
PROTON <25GeV										
Dubna	73	116	64	40	31	37	66	47	44	34
Moscow	10	50	25	23	31	42	58	17	34	26
Berkeley	909	780	497	517	353	216	61	57	25	16
Argonne ZGS	–	54	248	399	485	467	360	405	312	113
Rutherford	–	102	176	139	97	123	60	51	93	37
ELECTRON >5GeV										
Yerevan ARUS	–	–	–	0	9	14	8	5	28	32
Cambridge CEA	33	146	198	169	105	47	15	–	–	–
Hamburg DESY	–	18	304	279	246	181	88	97	14	3
SLAC	–	–	95	457	671	664	466	420	328	129
Daresbury NINA	–	–	–	19	73	71	41	19	15	13
Cornell	–	–	–	87	72	105	104	82	48	9
LINEAR e^- <5GeV										
Kharkov	–	–	5	18	31	71	25	21	19	29
Stanford Mk III	119	178	220	113	64	16	9	6	–	–
Orsay	17	40	36	15	7	13	6	–	–	–
e^+e^- COLLIDERS										
Novosibirsk	–	–	40	27	29	31	2	6	10	15
Orsay ACO + DCI	–	–	16	119	107	58	24	21	34	42
Frascati ADONE	–	–	–	5	63	170	231	49	45	28
Stanford SPEAR	–	–	–	–	–	6	866	826	363	304
Hamburg DORIS	–	–	–	–	–	0	90	357	462	115

Note: These figures were obtained by manual scanning of the *Science Citation Index*.

bias, their impact on the advance of knowledge in high-energy physics seems, in general, to have been uniformly low compared with their Western equivalents. Such a conclusion is not, however, the only possible one: for example, citation totals may be largely determined by the scale of research activity at each accelerator, and because this is generally smaller in the Eastern bloc, this may be the reason for the lower citation figures. To examine this possibility, we have calculated the average number of citations per paper (the 'citation rate') for each accelerator.[52] These figures are given in Table 5.

Again, the figures in Table 5 may be biased against Eastern-bloc accelerators. Of the three sources of bias discussed for Table 4, the second (papers omitted from our publication list) is obviously unimportant, since it merely reduces the sample size for averaging.[53] As we saw earlier, the total estimated bias from the other two sources

TABLE 5
Average Number of Citations Per Paper for Journal Articles Published During the Preceding Four Years

	1964	1966	1968	1970	1972	1974	1976	1978	1980	1982
PROTON >25GeV										
Serpukhov	–	–	–	4.8	3.0	2.8	2.1	1.2	1.2	1.3
	–	–	–	(10.7)	(6.4)	(4.6)	(3.9)	(2.6)	(2.3)	–
CERN PS	1.9	3.5	3.1	2.6	2.3	2.7	2.2	2.2	2.2	1.4
Brookhaven AGS	8.0	6.8	4.6	3.5	2.9	2.5	3.2	2.7	1.6	1.4
CERN ISR	–	–	–	–	12.9	14.0	7.4	5.4	4.4	4.7
Fermilab	–	–	–	–	3.8	6.9	6.7	7.3	3.7	2.7
CERN SPS	–	–	–	–	–	–	–	12.7	5.2	3.8
PROTON <25GeV										
Dubna	0.5	1.1	0.9	0.6	0.6	0.9	1.3	0.7	0.7	0.8
	(0.8)	(1.3)	(1.8)	(1.2)	(1.0)	(2.3)	(1.9)	(2.1)	(2.0)	–
Moscow	0.5	1.7	0.6	0.5	0.6	0.6	0.7	0.3	0.7	0.7
	(2.0)	(2.5)	(1.0)	(0.6)	(3.5)	(1.0)	(1.0)	–	–	–
Berkeley	3.4	3.8	2.8	2.9	2.5	2.3	1.3	2.5	1.7	2.0
Argonne ZGS	–	4.2	3.6	2.8	2.6	2.3	2.1	3.2	3.6	2.3
Rutherford	–	6.0	4.6	3.2	1.9	2.1	1.1	1.2	2.2	1.5
ELECTRON >5GeV										
Yerevan ARUS	–	–	–	–	0.8	0.8	0.6	0.3	1.4	1.5
Cambridge CEA	3.0	5.4	3.9	3.0	3.3	5.9	2.3	–	–	–
Hamburg DESY	–	1.6	6.1	4.4	4.0	2.6	1.6	2.5	0.7	0.6
SLAC	–	–	4.5	6.1	4.9	4.3	3.0	2.9	3.2	2.7
Daresbury NINA	–	–	–	2.1	2.7	2.3	1.8	1.0	0.7	0.8
Cornell	–	–	–	6.2	2.3	3.4	3.3	2.3	1.8	0.8
LINEAR e^- <5GeV										
Kharkov	–	–	0.8	0.7	0.6	1.3	0.5	0.4	0.5	0.8
Stanford Mk III	3.6	4.3	5.5	3.8	3.5	2.0	–	–	–	–
Orsay	1.3	1.8	1.6	0.9	0.8	1.9	0.9	–	–	–
e^+e^- COLLIDERS										
Novosibirsk	–	–	13.3	6.8	2.9	3.1	–	1.0	1.0	1.1
	–	–	–	–	(2.8)	(3.6)	–	(2.5)	(1.3)	–
Orsay ACO + DCI	–	–	–	14.9	9.7	7.3	3.0	3.5	4.3	2.8
Frascati ADONE	–	–	–	1.0	3.0	4.0	5.4	1.7	2.0	1.8
Stanford SPEAR	–	–	–	–	–	–	25.5	12.5	6.4	6.6
Hamburg DORIS	–	–	–	–	–	–	8.2	8.9	10.2	4.4

Note: These statistics are based on the data contained in Tables 3 and 4. The figures in brackets refer to the average citation rates for just the articles published in Western journals.

BEST COPY AVAILABLE

is probably in the range between 20 and 40 percent. However, the bias will only be this large for papers published in Eastern-bloc journals; for papers published in Western journals, it should be very much smaller.

Table 5 shows that papers from Serpukhov initially achieved quite a high citation rate — higher, for example, than those from early work at the CERN PS (the scale of research activity at the PS was, however, very much greater, as can be seen from Table 3). Indeed, for papers published in Western journals, the average Serpukhov citation rate came from close to the initial rate for the CERN ISR and SPS, appreciably ahead of Fermilab and even the Brookhaven AGS. It may be that the 'better' one-third papers reporting early Serpukhov results were channelled towards Western journals, because physicists from Europe and the US were involved in many of the early experiments. In contrast, the average impact of the remaining two-thirds of Serpukhov papers appears to have been very low indeed. As a result, apart from a brief period early in the 1970s, papers from the CERN PS and the Brookhaven AGS seem to have had a greater average impact than those from Serpukhov, despite the latter's considerable advantage in terms of energy and more recent construction.

Of the other Eastern-bloc accelerators, papers from Dubna, Moscow, Yerevan and Kharkov all consistently achieved a comparatively low impact, even after allowing for a bias of up to 40 percent. Only the earliest papers from Novosibirsk had a relatively high average impact, presumably because that collider was briefly the largest such facility in the world, and (in contrast with Dubna, which also temporarily held a world lead) its users exploited this advantage by carrying out at least a few novel and important experiments. However, this lead was short-lived, and, as Table 3 demonstrates, only a very small number of papers was published during this period. Overall, therefore, even after allowing for possible bias and the smaller scale of research activity, we must conclude that the average impact of publications from these machines seems to have been low in world terms.[54]

Highly-Cited Papers and Discoveries

A major criticism of all the output indicators discussed so far is that, while they provide comparative data on the total output of papers (most of which, at best, add only a small increment to the sum of scientific knowledge), they do not necessarily reveal which accelerators have been responsible for the occasional crucial or even 'revolutionary' discovery. However, such discoveries — and also slightly lower-level but nevertheless important 'advances' — can generally be identified from data on highly cited papers.[55] Out of our list of approximately 9000 articles published between 1961 and 1982, the top 2 percent earned 30 or more citations in a year. (The corresponding figures for the top 8 percent, 0.5 percent, and 12.5 percent are 15, 50 and 100 citations respectively.) Data on these highly-cited papers are given in Table 6.

We should again note the likely bias in these figures. There are two possible sources: from papers excluded from our list; and from papers which are included, but which 'lose' some of their citations. The first is likely to have a negligible effect, since the more important experimental results are almost without exception published in one of the major international journals.[56] However, the second is far from negligible — as much as 30 percent for the more international of the Eastern-bloc facilities (Serpukhov and Dubna), and rather more than this for the others. The two right-hand columns of Table 6 present the 'crude' data on highly cited papers adjusted to take account of biases of 20 and 40 percent.

In its early years, the Serpukhov accelerator produced a relatively large number of highly cited papers. Assuming a bias of approximately 20 or 30 percent, its record over the period 1969–72 is comparable with that of any other major accelerator: indeed, it was responsible for the only paper cited over one hundred times in a year.[57] Since then, however, it has yielded a negligible number of highly cited papers, far fewer even than the earlier-vintage and lower-energy CERN PS and Brookhaven AGS accelerators, let alone the new machines at Fermilab and CERN. As for the other Eastern-bloc accelerators, their record is relatively poor, even after adjusting for a bias of 40 percent. Perhaps the best that can be said is that some of them have come close to equalling the record of the less successful of the main Western facilities.

Peer-Evaluation

Our previous studies have suggested that, while indicators based on publication and citation counts provide essential information on

TABLE 6

Numbers of Highly Cited Experimental Papers for the Period 1961–80; Number of Papers Cited *n* or More Times in One Year

	1961–64				1965–68				1969–72				1973–76			
	n ≥ 15	*n* ≥ 30	*n* ≥ 50	*n* ≥ 100	*n* ≥ 15	*n* ≥ 30	*n* ≥ 50	*n* ≥ 100	*n* ≥ 15	*n* ≥ 30	*n* ≥ 50	*n* ≥ 100	*n* ≥ 15	*n* ≥ 30	*n* ≥ 50	*n* ≥ 100
PROTON >25GeV																
Serpukhov	–	–	–	–	–	–	–	–	12	4	2	1	7	0	0	0
CERN PS	9	1	0	0	46	1	0	0	19	1	0	0	18	4	2	0
Brookhaven AGS	24	5	1	1	42	12	1	0	22	1	0	0	15	5	2	1
CERN ISR	–	–	–	–	–	–	–	–	13	4	2	0	28	12	2	1
Fermilab	–	–	–	–	–	–	–	–	4	3	3	0	71	26	9	0
CERN SPS	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
PROTON <25GeV																
Dubna	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
Moscow	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Berkeley	33	5	2	0	9	0	0	0	8	1	0	0	2	0	0	0
Argonne ZGS	–	–	–	–	6	0	0	0	7	0	0	0	5	0	0	0
Rutherford	–	–	–	–	9	0	0	0	0	0	0	0	0	0	0	0
ELECTRON >5GEV																
Yerevan ARUS	–	–	–	–	–	–	–	–	0	0	0	0	0	0	0	0
Cambridge CEA	1	0	0	0	7	0	0	0	2	0	0	0	0	0	0	0
Hamburg DESY	–	–	–	–	10	1	0	0	3	0	0	0	0	0	0	0
SLAC	–	–	–	–	4	1	0	0	21	4	2	0	14	2	1	0
Daresbury NINA	–	–	–	–	–	–	–	–	0	0	0	0	0	0	0	0
Cornell	–	–	–	–	–	–	–	–	4	0	0	0	1	0	0	0
LINEAR e^- <5GeV																
Kharkov	–	–	–	–	0	0	0	0	0	0	0	0	0	0	0	0
Stanford Mk III	7	1	0	0	8	1	0	0	2	1	0	0	0	0	0	0
Orsay	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
e^+e^- COLLIDERS																
Novosibirsk	–	–	–	–	1	1	0	0	0	0	0	0	0	0	0	0
Orsay ACO + DCI	–	–	–	–	1	0	0	0	4	1	0	0	0	0	0	0
Frascati ADONE	–	–	–	–	–	–	–	–	2	1	0	0	3	1	1	1
Stanford SPEAR	–	–	–	–	–	–	–	–	0	0	0	0	23	12	7	5
Hamburg DORIS	–	–	–	–	–	–	–	–	0	0	0	0	3	2	1	0

Note: These figures were obtained by manual scanning of the *Science Citation Index*.

298

TABLE 6 (continued)

	1977–80				1961–80 figures adjusted for 1961–80				1961–80 figures adjusted for 20% bias against E. Europe				40% bias against E. Europe			
	$n \geq$ 15	$n \geq$ 30	$n \geq$ 50	$n \geq$ 100	$n \geq$ 15	$n \geq$ 30	$n \geq$ 50	$n \geq$ 100	$n \geq$ 12	$n \geq$ 24	$n \geq$ 40	$n \geq$ 80	$n \geq$ 9	$n \geq$ 18	$n \geq$ 30	$n \geq$ 60
PROTON >25GeV																
Serpukhov	0	0	0	0	19	4	2	1	32	6	2	1	49	15	4	1
CERN PS13	13	2	1	0	105	9	3	0								
Brookhaven AGS	0	0	0	0	103	23	4	2								
CERN ISR	11	2	0	0	52	18	4	1								
Fermilab	40	10	5	1	115	39	17	1								
CERN SPS	19	7	3	0	19	7	3	0								
PROTON <25GeV																
Dubna	0	0	0	0	2	0	0	0	3	0	0	0	8	1	0	0
Moscow	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Berkeley	0	0	0	0	52	6	2	0								
Argonne ZGS	8	3	0	0	26	3	0	0								
Rutherford	1	0	0	0	10	0	0	0								
ELECTRON >5GEV																
Yerevan ARUS	1	0	0	0	1	0	0	0	1	0	0	0	2	0	0	0
Cambridge CEA	–	–	–	–	10	0	0	0								
Hamburg DESY	0	0	0	0	13	1	0	0								
SLAC	7	3	1	1	46	10	4	1								
Daresbury NINA	0	0	0	0	0	0	0	0								
Cornell	0	0	0	0	5	0	0	0								
LINEAR e^- <5GeV																
Kharkov	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Stanford Mk III	–	–	–	–	17	3	0	0								
Orsay	–	–	–	–	0	0	0	0								
e^+e^- COLLIDERS																
Novosibirsk	0	0	0	0	1	1	0	0	2	1	0	0	4	1	0	0 0
Orsay ACO + DCI	0	0	0	0	5	1	0	0								
Frascati ADONE	1	0	0	0	6	2	1	1								
Stanford SPEAR	19	3	0	0	42	15	7	5								
Hamburg DORIS	16	7	2	0	19	9	3	0								

relative contributions to scientific progress, they must always be considered alongside peer-evaluation data. This conclusion probably applies even more strongly to East–West comparisons where, as we have seen, the bibliometric data are significantly biased. How, then, in the opinion of high-energy physicists in the East and West, do the scientific outputs from Eastern-bloc accelerators compare with those from equivalent facilities in Western Europe and the US?

We obtained peer-evaluation data for all the accelerators in the present study by interviewing 182[58] experimental and theoretical high-energy physicists during late 1981 and early 1982. Our sample consists of 52 physicists in the US, 112 from six countries in Western Europe (including a number who had worked on experiments in the Eastern bloc), and 18 from groups in three East European countries regularly using accelerators in the Soviet Union. The interviews were intensive (typically lasting 1 ½ to 2 hours) and structured, being based on a common set of questions but with some additional questions for special groups of researchers. After giving brief details of their background and career, interviewees were asked first to describe their own research and perceived contributions to high-energy physics; next, to identify the principal contributions of the various collaborations in which they had worked; then those from the accelerators they had used; and, finally, the overall contributions from the world's other major accelerator facilities. This sequence helped respondents prepare themselves for the peer-evaluation section of the interview. In this, they were invited to rank the accelerators at fourteen laboratories in order, according to their overall contributions to high-energy physics during the period from 1969 to 1978. Relatively few were able to rank all fourteen in full order from first to fourteenth: most regarded at least two as having made equivalent contributions, particularly among those lower in the order; others preferred to identify five or six distinct groups of accelerators, and then to rank these in order. Only eight percent found the ranking too difficult, or refused for personal or professional reasons. We then calculated average rankings on a scale of 1 (top) to 14 (bottom);[59] the results are given in Table 7. (The final column of the Table converts the overall rankings back into the relative positions of the accelerators at the fourteen laboratories.)

Although there was a slight tendency for Americans to rank their own accelerators rather more highly than did West Europeans, and vice versa, the rankings by these two groups are, overall, very

TABLE 7

Peer-Evaluation Rankings of the Contributions to High-Energy Physics Made by Accelerators at 14 Laboratories between 1969 and 1978

	Accelerator(s)	Rankings[a] by high-energy physicists in:			Self-ranking (sample size)	Peer-ranking (sample size)	Overall average rankings (n=168)	Average relative position
		US (n=49)[b]	W. Europe (n=105)	E. Europe (n=14)				
PROTON >25GeV	Serpukhov	9.0(±0.3)	8.8(±0.2)	7.1(±0.6)	6.7(±0.4) (n=19)	9.0(±0.2) (n=149)	8.7(±0.2)	7=
	CERN (PS, ISR, SPS)	2.6(±0.1)	2.4(±0.1)	2.0(±0.2)	2.3(±0.1) (n=108)	2.6(±0.1) (n=60)	2.4(±0.1)	2
	Brookhaven AGS	3.6(±0.2)	3.7(±0.1)	3.6(±0.3)	3.5(±0.2) (n=47)	3.7(±0.1) (n=121)	3.7(±0.1)	4
	Fermilab	2.8(±0.1)	3.3(±0.1)	3.0(±0.3)	2.6(±0.1) (n=45)	3.3(±0.1) (n=123)	3.1(±0.1)	3
PROTON <25GeV	Dubna	12.0(±0.2)	11.5(±0.1)	10.2(± 0.5)	11.2(±0.4) (n=9)	11.5(±0.1) (n=159)	11.5(±0.1)	12
	Moscow	12.5(±0.2)	12.1(±0.1)	12.2(±0.3)	–	12.2(±0.1) (n=168)	12.2(±0.1)	13=
	Argonne ZGS	6.8(±0.2)	8.0(±0.2)	7.8(±0.5)	7.5(±0.5) (n=12)	7.6(±0.1) (n=156)	7.6(±0.1)	6
	Rutherford	9.2(±0.2)	9.0(±0.2)	10.1(±0.5)	7.5(±0.4) (n=23)	9.4(±0.1) (n=145)	9.1(±0.1)	9=
ELECTRON	Yerevan ARUS	12.3(±0.2)	12.2(±0.1)	11.9(±0.4)	–	12.2(±0.1) (n=168)	12.2(±0.1)	13=
	Cambridge CEA	8.9(±0.3)	9.3(±0.2)	10.2(±0.4)	–	9.2(±0.1) (n=165)	9.2(±0.1)	9=
	DESY (7GeV, DORIS)	5.7(±0.2)	5.3(±0.1)	5.4(±0.3)	5.1(±0.4) (n=17)	5.4(±0.1) (n=151)	5.4(±0.1)	5
	SLAC (20GeV, SPEAR)	1.3(±0.1)	1.2(±0.1)	1.6(±0.2)	1.1(±0.1) (n=17)	1.3(±0.1) (n=151)	1.3(±0.1)	1
	Daresbury NINA	10.1(±0.2)	9.8(±0.1)	10.7(±0.5)	–	10.0(±0.1) (n=163)	10.0(±0.1)	11
	Cornell	8.2(±0.2)	8.6(±0.2)	9.1(±0.5)	–	8.5(±0.1) (n=165)	8.5(±0.1)	7=

[a] These rankings are based on the results of interviews carried out in 1981 and 1982. [b] *n* = sample size.

similar. The Western rankings are also fairly consistent with those made by Eastern-bloc physicists, with two exceptions — while the latter gave virtually the same rankings to the Moscow and Yerevan accelerators as did Westerners, they ranked somewhat higher the work at Dubna and Serpukhov. Either East Europeans hold inflated views on the past performance of these two machines; or Western physicists are unaware of some of their contributions. Let us consider the first possibility. In this survey, as in others we have carried out, there is a small but significant tendency for researchers to over-rate the significance and contributions from research facilities they have used. Thus, for example, as Table 7 shows, physicists who have used the Fermilab accelerator gave it a 'self-ranking' of 2.6 – significantly better than the 'peer-ranking' of 3.3 given by those who have done their experimental work elsewhere. However, it seems doubtful whether this can explain all the variation between Eastern and Western rankings of Serpukhov and Dubna; at least part is likely to be due to an 'ignorance' effect among Western physicists. However, even if one allowed for this, it would only raise Serpukhov from a position to seventh equal with the Cornell accelerator to sixth equal with Argonne, well behind the main Western proton accelerators; and the Dubna machine would only rise from twelfth to eleventh equal with Daresbury, still behind Argonne and Rutherford (and almost certainly the Berkeley Bevatron, if it had been included in the list). As for Moscow and Yerevan, physicists in both East and West agreed that their scientific records over the period 1969–78 put them at the bottom of this list of the world's highest-energy accelerators.[60]

However, the peer-ranking data in Table 7 only give a rather general picture of the output from the accelerators at the fourteen laboratories. Since certain laboratories, such as CERN and SLAC, operate several accelerators, it was necessary to obtain more specific information on scientists' perceptions of the relative performance of *individual* accelerators. The data also only represent the judgements of scientists who have attempted, for each accelerator laboratory, to weigh up the relatively small number of discoveries or major advances for which it has been responsible with the very much greater number of lower-level contributions.[61] Such general rankings are not without utility, but more specific data are also required. We therefore focused in detail upon six of the world's principal proton facilities: the CERN ISR, PS and SPS, the Brookhaven AGS; and the Fermilab and Serpukhov

[illegible] AVAILABLE

synchrotrons. High-energy physicists were asked to assess these on a ten-point scale (10 = most successful), distinguishing between their records in making 'discoveries' and in producing 'more precise measurements'. The results are given in Table 8. For 'discoveries', Brookhaven was consistently placed ahead of the other five, while the CERN PS was judged the most successful in terms of 'more precise measurements'. Again differences appear between the views of Eastern-bloc researchers and those of Western

TABLE 8

Assessments (on a 10-point scale[a]) of main proton accelerators in terms of (a) 'discoveries' (b) providing more precise measurements

		Assessments by High-Energy Physicists in:			Self-Assessment	Peer-Assessment	Overall Average Assessments
		United States	Western Europe	Eastern Europe			
Discoveries	IHEP Serpukhov (76GeV)	2.1(±0.2)[b] (n=50)[c]	2.8(±0.2) (n=104)	4.2(±0.5) (n=15)	3.8(±0.5) (n=20)	2.6(±0.1) (n=149)	2.7(±0.1) (n=169)
	CERN PS (28GeV)	6.0(±0.3) (n=50)	7.2(±0.2) (n=104)	7.7(±0.4) (n=15)	7.1(±0.2) (n=86)	6.7(±0.2) (n=83)	6.9(±0.1) (n=169)
	Brookhaven AGS (33GeV)	9.6(±0.1) (n=50)	9.0(±0.1) (n=104)	8.6(±0.5) (n=15)	9.5(±0.1) (n=48)	9.0(±0.1) (n=121)	9.2(±0.1) (n=169)
	CERN ISR (31+31GeV)	5.8(±0.3) (n=50)	5.8(±0.2) (n=104)	8.4(±0.3) (n=15)	6.8(±0.3) (n=36)	5.9(±0.2) (n=133)	6.1(±0.2) (n=169)
	Fermilab (500GeV)	7.1(±0.2) (n=50)	7.1(±0.2) (n=104)	8.3(±0.3) (n=15)	7.4(±0.3) (n=46)	7.1(±0.1) (n=123)	7.2(±0.1) (n=169)
	CERN SPS (500GeV)	5.3(±0.2) (n=50)	5.8(±0.2) (n=104)	6.8(±0.4) (n=15)	5.9(±0.3) (n=68)	5.6(±0.2) (n=101)	5.7(±0.1) (n=169)
More precise measurements	Serpukhov	2.8(±0.2) (n=49)	3.7(±0.2) (n=103)	4.9(±0.6) (n=15)	4.3(±0.5) (n=20)	3.5(±0.2) (n=147)	3.6(±0.2) (n=167)
	CERN PS	8.3(±0.2) (n=49)	8.6(±0.1) (n=103)	8.7(±0.4) (n=15)	8.5(±0.1) (n=86)	8.5(±0.1) (n=81)	8.5(±0.1) (n=167)
	Brookhaven AGS	7.4(±0.2) (n=49)	7.0(±0.2) (n=103)	8.1(±0.4) (n=15)	7.1(±0.2) (n=47)	7.2(±0.2) (n=120)	7.2(±0.1) (n=167)
	CERN ISR	6.5(±0.3) (n=49)	7.0(±0.2) (n=103)	8.3(±0.3) (n=15)	7.3(±0.3) (n=35)	6.9(±0.2) (n=132)	7.0(±0.1) (n=167)
	Fermilab	6.1(±0.2) (n=49)	5.8(±0.2) (n=103)	7.9(±0.3) (n=15)	6.3(±0.2) (n=46)	6.0(±0.2) (n=121)	6.1(±0.1) (n=167)
	CERN SPS	8.0±0.2) (n=49)	8.3(±0.1) (n=103)	8.0(±0.3) (n=15)	8.2(±0.2) (n=67)	8.2(±0.2) (n=100)	8.2(±0.1) (n=167)

a. 10 = most successful. The assessments are based on the judgements of high-energy physicists of the relative scientific outputs from the accelerators over their entire period of operation up to the time of the interviews in late 1981/early 1982.

b. The figures in brackets indicate the root-mean-square variations between the assessments made by the different groups of high-energy physicists, giving some approximate idea of the divergence of opinion within each group.

c. Sample size.

Europeans and Americans on the scientific performance of the Serpukhov accelerator. However, even the Eastern-bloc physicists placed Serpukhov considerably behind all five Western facilities in both rankings.

An Overall Assessment

We stressed earlier that, if reliable conclusions are to be drawn about the relative scientific performance of Eastern and Western accelerators, there must be some convergence between the different partial indicators used in the assessment. To what extent, then, do the peer-evaluation results accord with those based on bibliometric data?

Taking first the Serpukhov accelerator, we have seen that its annual output of experimental publications was less than that of all the major Western proton machines with the exception of the CERN ISR;[62] and, even after allowing for any bias, those papers have generally earned less citations. This strongly suggests that the overall impact of Serpukhov work on the advance of knowledge has been less than that of its main Western competitors; confirmatory evidence for this comes from the peer-evaluation assessment using the criterion of 'more precise measurements'. As for 'discoveries', even assuming a citation bias of up to 30 percent, the figures still reveal that Serpukhov produced slightly fewer highly cited papers than the ISR, and considerably less than the other major Western proton machines;[63] this is again consistent with the peer-evaluation results.[64] The conclusion is that users of the Soviet accelerator failed to capitalize fully upon its position between 1967 and 1972 as the world's highest-energy accelerator — in marked contrast to users of the Brookhaven AGS, and to a lesser extent the CERN PS, who exploited new machines that, between 1960 and 1967, held a similar comparative advantage — as, indeed, did Fermilab from 1972 to 1976.

Similar, but even heavier, criticism can be levelled against the Dubna accelerator. This held a world energy lead from 1957 to the end of 1959, and it might have been expected to remain a front-line machine through the 1960s. The publication and citation records of machines of roughly similar energy (Berkeley, Argonne and Rutherford) show that, even after 1960, when higher-energy facilities came into operation at CERN and Brookhaven, much

BEST COPY AVAILABLE

important experimental work could still be carried out in this energy region. According to our data, however, Dubna users could not take advantage of its energy;[65] even allowing for bias, Dubna's publication and citation record is appreciably worse than Berkeley and Argonne, and also worse than the slightly smaller Rutherford accelerator. Furthermore, this conclusion is evidently shared by the large majority of our interviewees, although they did place Dubna slightly ahead of the accelerators at Moscow and Yerevan — again consistent with their bibliometric records.

The Kharkov and Novosibirsk accelerators were not included in the peer-ranking exercise, and so we have no systematic peer-evaluation data for them. Nevertheless, we obtained a range of qualitative data, in particular about the output of Novosibirsk; a number of physicists cited the centre's early results on electron-positron collisions as significant during the mid-1960s (which is consistent with the data on highly-cited papers). There was also a widespread view among physicists, in both East and West, that Novosibirsk had contributed significantly to the development of accelerator physics and technology, and was rated highly in terms of this criterion even compared with most large Western laboratories.[66] As for the Kharkov linear accelerator, its experimental work seems to have had virtually no impact outside the Soviet Union.

In summary, although the various partial indicators used in this study are without doubt biased against the Eastern bloc, we can estimate the approximate magnitude of these biases and adjust the output indicators accordingly. The adjusted partial indicators based on publications, citation counts, highly-cited papers and peer-evaluation do appear to yield reasonably convergent results. Certain conclusions can then be drawn about the relative contributions to scientific progress made by the various Eastern-bloc accelerators. In each case, the contributions have been generally rather small when compared with equivalent Western facilities. While the more successful accelerators have perhaps come close to matching the scientific performance of the least successful of their Western competitors, others have not managed even this. One East European physicist aptly summarized this conclusion:

> Scientists in both the East and the West have the same opinions and views concerning the high energy physics labs where the most important scientific results and innovations have been made. These are all in the West. (Interview, 1982)[67]

Let us now analyze some possible reasons for this difference in scientific performance between East and West.

Factors Affecting the Scientific Performance of Eastern-Bloc Accelerators

We adopted two approaches to identify and analyze the factors responsible for the difference in scientific performance between East and West. First, high-energy physicists, after ranking Serpukhov and the other main accelerators in terms of 'discoveries' and 'more precise measurements', were questioned about possible reasons for differences in performance. We classified the content of the responses in terms of ten general categories (these are obviously to some extent related). The results are summarized in Table 9. Second, at the end of each interview, we asked physicists to complete an attitude survey containing approximately thirty statements on many of the issues discussed in the interview. Some of these statements were rather controversial, so we have to provide a fairly wide set of response categories, using a 7-point scale. Thus, for each statement, interviewees had the option of (1) 'agreeing strongly', (2) 'agreeing', (3) 'agreeing with reservations', (4) 'being neutral or holding mixed views', and so on, up to (7) 'disagreeing strongly'. While the first approach produces qualitatively better information, the statistics yielded by the content analysis give only a minimum limit for the percentage of interviewees believing each factor to have been important (see note to Table 9). The second approach is much better from the latter point of view, and we use it at various points in the text to support our conclusions.

For example, the first statement in the attitude survey was aimed at establishing the degree of satisfaction (or dissatisfaction) felt by interviewees towards the research facilities at the laboratories where they had carried out experiments. The statement was of the following form:

> The system for providing experimental facilities for high-energy physicists at CERN/at Brookhaven/at Fermilab/in Eastern-bloc laboratories is working so well that it does not need changing.

The responses suggest that CERN users were most satisfied with the research facilities available to them; 60 percent agreed with the

BEST COPY AVAILABLE

TABLE 9

Factors Explaining the Relative Scientific Performance of the Serpukhov Accelerator (percentage of interviewees believing this factor to have been important)

	Users of Serpukhov Accelerator (n = 20)	Others (n = 93)	All Interviewees (n = 113)
1 Poor/inadequate detectors — e.g. as a result of lack of fast electronics	65%	68%	67%
2 Poor computer facilities	55%	28%	33%
3 Accelerator poorly designed (old-fashioned) and inadequate performance (e.g. poor beams)	35%	17%	20%
4 Accelerator energy not high enough – only a factor of two increase on previous machines . . . soon overtaken	20%	16%	17%
5 Insufficient resources (e.g. as a result of industrial supply problems) and technical support	60%	45%	48%
6 Poor scientific management at Serpukhov	30%	30%	30%
7 Poor contacts between research groups in Eastern Europe, and between East and West	35%	26%	27%
8 Poor competitive ethos in Eastern Europe — little scope for individual initiative, motivation problems, and so on	25%	23%	23%
9 Researchers have less experimental experience than in the West	5%	16%	14%
10 General bureaucratic problems of Eastern European scientific system — e.g. inflexible five-year plans	56%	57%	58%

Note: These figures represent minimum values only, since they are based on a content analysis of answers to a general question concerning the factors structuring Serpukhov's scientific performance. If the interviewees had instead been taken systematically through the above list of factors, we would undoubtedly have obtained considerably higher figures for the percentage believing each factor to have been important.

BEST COPY AVAILABLE

statement, while only 36 percent disagreed. Fermilab users were also relatively satisfied — the corresponding figures were 54 percent and 37 percent respectively. The situation was rather different for Brookhaven, where major reservations and criticisms about the management of the ISABELLE project (to build a large proton-proton collider) were reflected in the results — only 19 percent agreed, while 70 percent disagreed. However, it was the users of the Eastern-bloc accelerators who appeared most dissatisfied — none agreed, while 94 percent disagreed with the relevant statement. So with what aspects of the facilities were they particularly dissatisfied? We will now focus mainly on Serpukhov since this is the main Soviet facility used by non-Soviet researchers, as well as the largest Eastern-bloc accelerator.

Table 9 shows that one of the main problems facing Serpukhov has concerned the provision of adequate detectors. Although the accelerator itself was officially completed in 1967, much of the subsidiary instrumentation and all the main detector facilities were not ready until several years later; by then the larger Fermilab accelerator was operating. The same situation had afflicted the CERN PS a few years earlier,[68] putting its experimental programme at an appreciable disadvantage to the Brookhaven AGS during the early 1960s. At Serpukhov, whereas the construction of the accelerator itself received priority in the national plan, so that adequate resources and highly-skilled technical personnel were made available, the equally important task of preparing instrumentation and detectors seems to have had lower priority. Responsibility for this cannot simply be placed on apparent rigidities in the planning system (see factor 10, below), making it more difficult to obtain priority for secondary equipment; some blame must also be attached to the scientific community. As one East European scientist commented,

> If you want to build a new machine, you must think in advance about the detectors and experimental apparatus. The big mistake we made was not to think in advance about how to use the machine effectively. The result, therefore, was that the machine has not been used effectively. (Interview, 1982)

Given that a similar situation had arisen ten years earlier at Dubna, it is perhaps surprising that this mistake was not anticipated and avoided. Moreover, as several researchers pointed out, the main detectors were not only late, but some at least (for example, the

Ludmilla and SKAT bubble-chambers) suffered various technical problems, and most had limited performance compared to their far more sophisticated Western equivalents. Only relatively simple (though important) first-generation experiments were therefore possible, leaving the more sophisticated experiments to be performed elsewhere. Even then, much of the early work of high impact was carried out using Western-built detectors, particularly the French bubble-chamber, Mirabelle.

Many of our interviewees felt that the key problem in developing sophisticated instrumentation is obtaining access to high-technology industry. Since that industry has met the requirements of the Soviet military and space programmes, it clearly *could* in principle meet the needs of experimental high-energy physicists. However, firms gain significant prestige by working on military and space projects: some work only on such projects, while others are willing to undertake technologically advanced contracts in other sectors because this allows them to fulfil quotas set by the national plan. High-energy physics has no such advantages. Except when resources are specifically marshalled to construct a large accelerator, particle physics does not have sufficient priority to warrant the special attention of high-technology industry. High-energy physicists, therefore, are not the most attractive of customers; trying to meet their very exacting demands is risky, and they order only small numbers of each item. As one researcher put it:

> Firms do not want to work for high-energy physics centres since there is not enough profit in it, as well as being demanding and risky work. The factories see it as more profitable to stick to mass-production goods. (Interview, 1982)

It is thus extremely important for researchers to have strong, day-to-day, personal links with high-technology firms to gain access to state-of-the-art equipment and techniques in, for example, fast electronics — yet this is something most Eastern-bloc scientific centres find it difficult to achieve.[69] In contrast, the Western market mechanism has produced small advanced-technology firms which specifically service the needs of accelerator centres.[70] Irrespective of whether the market mechanism is the most appropriate way of determining industrial priorities, it does offer an explanation for the differing fortunes of experimental high-energy physics in the East and West.

The same industrial supply problems probably account for the second factor cited as limiting Serpukhov's scientific performance —

its computing facilities. Computers are necessary not only for the control and automatic operation of the accelerator (especially for monitoring the particle beam), but also for the transfer and processing of data. Initially, the Soviet-made computers at Serpukhov had rather limited capacity, and were, by all accounts, unreliable. The situation improved early in the 1970s when, after high-level discussions between the UK and US governments, authorization was given for the sale of large ICL mainframe computers. However, the computing capacity at Serpukhov remained limited as experimental activity increased, while servicing and using these foreign computers efficiently have led to problems (for example, authorization is needed for foreign-currency purchases of spare parts or additional equipment). Moreover, although the central computing facilities at Serpukhov have been considerably improved, most Eastern-bloc user-groups have rather poor home facilities, with the result that researchers can take weeks to complete data-processing that would take hours in most Western universities.

A third factor is the intrinsic quality of the Serpukhov accelerator itself.[71] Its design represented an extrapolation from the earlier CERN PS and Brookhaven machines, and was therefore not particularly innovative in world terms. The resulting accelerator has a poor repetition-rate and produces beams of relatively low intensity, two factors which have greatly constrained its experimental output; and, because of the computer problems, interviewees reported that the beam was neither very accurate nor especially well controlled. There have also been difficulties with the external beams, especially the neutrino beam. Originally, it had been hoped to avoid at least some of the technical problems anticipated with accelerator control and detector construction by encouraging the participation of Western scientists in experiments at Serpukhov, bringing with them, *inter alia*, some of the modern technology not readily available to high-energy physicists in Eastern Europe. However, the accelerator's inherent limitations soon dampened the interest of Western scientists, despite the significant Western investment in equipment for Serpukhov, by CERN and the French Government in particular. Most of the Western users of Serpukhov whom we interviewed reported being disillusioned by the accelerator's quality during these crucial early years when it was the world leader.

Also contributing to this fairly rapid loss of interest was the fact

BEST COPY AVAILABLE

that the accelerator's energy represented a rather modest increase on that of previous machines (see factor 4 in Table 9). This relatively narrow, new energy range proved not to be very exciting compared with that opened up by the Brookhaven AGS and CERN PS accelerators, which in 1960 had had a four- to five-fold energy advantage over previous machines. By the time the major Serpukhov detectors were operational, the Fermilab accelerator, with its higher energy and more advanced experimental facilities, was coming into operation, as was the CERN ISR. Many Western high-energy physicists therefore inevitably migrated from Serpukhov to prepare for experiments elsewhere.

Whereas factors 1 to 4 relate essentially to problems with the Serpukhov accelerator and its hardware, factors 5 and 6 are concerned more with the way in which the facility has been supported and operated. Although we have no reliable figures on the funding levels of Eastern-bloc high-energy laboratories, their annual operating budgets seem to be much lower than for equivalent Western facilities. Of those interviewed, 88 percent agreed with the statement, 'Soviet accelerator centres have not been given sufficient technical resources to enable them to compete successfully with their Western counterparts', while none disagreed. In the opinion of many (see factor 5 in Table 9), national plans do not, in general, allocate to high-energy physics the financial and technical resources it requires. Furthermore there is a shortage of human resources. In Western Europe and the USA, a highly developed and extensive community of users, based largely in universities, carry out experiments at the major accelerator centres. Apart from Serpukhov and Dubna, Soviet accelerators tend, for a mixture of institutional and geographical reasons,[72] to have only a very restricted user-community, and are not so intensively used as Western facilities.

A number of other scientists (30 percent — see factor 6) criticized the senior scientific management at Serpukhov, comparing it unfavourably with that of large Western laboratories. This problem was, in turn, seen in a wider socio-political context.[73] Interviewees were particularly critical of what they saw as the overly hierarchical structure of Eastern-bloc science, with strategic decisions being made by a small and rather elderly élite, whose promotion is not always for entirely scientific reasons. Running a large research institute involves significant administrative and political responsibilities, leaving little time for active involvement in research. Given that there appears to be little inter-institute mobility, and no formal retirement age for directors, it is perhaps not surprising that many interviewees claimed that senior management does not always retain a firm grasp of recent research developments. Such leaders cannot be in the strongest position to plan an experimental programme to exploit a new research facility, nor to decide which of the experiments proposed should be given priority. Of the East Europeans interviewed, three times as many agreed as disagreed with the statement: 'The system by which the allocation of time on Soviet accelerators is decided tends to encourage too much routine research rather than highly innovative but risky experiments'. Some argued further that the concentration of decision-making in the hands of a small administrative élite had a stultifying effect, one casualty of which was the motivation of researchers. This was seen as generally weaker than in the West, where more aspects of scientific decision-making are decentralized, and greater opportunities consequently exist for scientists to participate in determining the direction of a laboratory's research activity.

So far, we have mainly been concerned with explaining the performance of the Serpukhov accelerator only.[74] However, Serpukhov is almost certainly accorded a higher level of priority and support than laboratories with accelerators of lower energy, where the problems discussed above are probably more, rather than less, severe.

Factors 7 to 9 refer to general problems faced by the whole experimental high-energy physics community in the Eastern bloc, rather than by the user-groups associated with one particular facility. The first concerns the relatively poor contacts both between research groups in the Eastern bloc, and between researchers in East and West. Some interviewees described the relationship between Eastern research groups as being far too secretive, rather than fruitfully collaborative, with sharing of skills and equipment. To some extent, this may reflect difficulties in telephoning, telexing, photocopying, and writing letters to, or visiting, foreign collaborators. For both Dubna and Serpukhov, it was argued that the resulting fragmentation of effort has seriously weakened the overall effectiveness of their research programmes.

However, the paucity of contacts with the Western high-energy physics community has probably had the more serious long-term repercussions. Not only are the possibilities of foreign travel

BEST COPY AVAILABLE

generally rather limited (except for the more senior scientists), but researchers are not well integrated into the informal communication system. This is partly due to restrictions on researchers circulating unpublished material to scientists outside their institute, particularly those in the West. Late in the 1960s, for example, Serpukhov researchers were slow to recognize the importance of Western developments on the discovery of 'scaling' and the formulation of the parton model.[78] If their significance had been appreciated earlier, Serpukhov workers might have embarked upon at least some experiments in this leading-edge of physics, and these could conceivably have had a very considerable impact early in the 1970s.

The Soviet Union is certainly not unaware of such constraints on its researchers, as the following interview quotation makes clear:

> I am convinced and have evidence that the question of opening up Serpukhov to the West was hotly debated within the Soviet administration, with strong views in favour and against being put forward. The argument that finally won was that the disastrous experience with the Dubna synchrophasotron could not be repeated. Western experience and technology were needed to ensure an adequate experimental output from what was for years the world's highest-energy accelerator. The counter-argument concerned the 'ideological contamination' of Soviet physicists. In the end it was decided the lesser evil was to open Serpukhov up. (Interview, 1982)

However, even when Western physicists did come to work at Serpukhov, the informal mixing with other researchers was, according to those involved, disappointing — much less, for example, than at CERN, where West European experimentalists have learned a vast amount from visiting foreign scientists. Several CERN visitors to Serpukhov cited social isolation as one of the main problems during their stay in the Soviet Union.

The user-community of Eastern-bloc accelerators was also seen as having a less developed competitive ethos than in the West. One interviewee described the situation in the East as follows:

> There is no competition: experiments are thought out and run by the 'bosses'. The result is that there is no motivation for young people to stand out and take the initiative. In fact, there is a general lack of criticism and debate because of the hierarchy and the wish not to offend the people at the top. It is particularly difficult for young people. (Interview, 1982)

While one must recognize the value-laden connotation of the term 'competition', as well as the negative effects upon individual scientists of the occasionally rather ruthless behaviour found in certain Western laboratories[76] (which several East European interviewees strongly condemned), it does appear that, in the West, young physicists have greater opportunities to devise their own experiments, bid for time on accelerators, and, if successful, lead research collaborations. With this freedom and the higher level of motivation it engenders, Western researchers are more likely to have been involved in a greater number and variety of experiments than their Eastern counterparts. The latter therefore tend to have a lower level of experimental experience (factor 9 in Table 9), and thus fail to gain the tacit knowledge that is so essential if, after an unforeseen result or opportunity, an experiment's direction has to be suddenly changed in mid-stream. According to some early workers at Serpukhov, Eastern-bloc researchers not only had a poor idea of the important things to look for (in particular, quarks and the parton-structure of hadrons), but also a rather limited repertoire of experimental techniques — adequate for carrying out the simpler and more obvious experiments on the Serpukhov accelerator, but not sufficient for tackling precise and ambitious second-generation experiments, most of which, as we have seen, were left to machines in the West.

The final factor we shall consider was one of the two most frequently cited — the problem of bureaucracy in Eastern-bloc science. For the statement: 'The major problem facing the users of Soviet accelerators is the over-bureaucratized system under which these facilities are operated', no less than 88 percent of our East European interviewees agreed to a greater or lesser extent, while none at all disagreed. One of the more critical stated:

> The administration and bureaucracy do not work. Partly the system is corrupt, partly it is malfunctioning, and partly no-one wants to take responsibility for anything in order to avoid being blamed for mistakes. (Interview, 1982)

This, of course, is a somewhat extreme view. (A number of Western physicists were also highly critical of certain aspects of their own administrative systems.) Nevertheless, interviewees gave many examples of the effects of bureaucracy: tasks that in the West would be carried out with little fuss in the East require much paperwork and numerous authorizing signatures;[77] the publication of research papers generally takes longer because of the need to obtain high-level authorization; activities involving several research institutes or

BEST COPY AVAILABLE

facilities (such as a programme of visits by a foreign scientist, or setting up a large collaboration) will typically require the agreement of three separate Ministries; an accelerator's experimental programme cannot generally be changed quickly (even if some unexpected crucial discovery demands that it should be made) since too many layers of bureaucracy have first to be negotiated. According to many interviewees, all this stems, in large part, from one root cause — the over-reliance on a centralized five-year planning system that has become rigid and inflexible. These comments are not untypical:

> The Russians have a five-year plan for everything, even high-energy physics . . . A five-year horizon is the time to ask for money for experiments, and things have to be planned rigidly that far in advance. Flexible re-allocation is not possible as in the West. (Interview, 1982)

Experiments planned years in advance are in some cases dutifully carried out, even if they have long since been rendered unnecessary by developments elsewhere. Conversely, experiments that could not have been foreseen earlier sometimes cannot be integrated into the planned experimental programme. For example, in 1971, high-energy physicists suddenly became excited at the possibility of integrating electromagnetic and weak interactions under one unified theory.[78] The candidate unified theory predicted a new and previously unobserved form of interaction, involving weak 'neutral currents'. Experimentalists immediately began to search for these phenomena, and they were discovered two years later on the CERN PS. If there had been greater flexibility at Serpukhov, the commissioning of the neutrino beam could have been speeded up. An experiment might then have been mounted with the French bubble-chamber, Mirabelle, that would have stood a good chance of being first to see the predicted effects. However, such a rapid shift was never feasible, and the discovery, one of the most important in the field during the 1970s, was instead made at CERN. Many other such examples demonstrate that flexible and decentralized strategies must be adopted in running large experimental facilities in an intensely competitive branch of science characterized by rapid shifts of interest and sudden unpredicted developments. Researchers who work in a scientific system based on hierarchical control, highly-centralized decision making, and relatively inflexible long-term plans, are obviously labouring under a

BEST COPY AVAILABLE

considerable disadvantage.[79] To improve the situation, considerable structural reforms appear to be necessary in the organization and administration of science — a task that a number of Eastern-bloc governments now seem to view as a priority.[80]

Summary and Conclusion

To conclude, we began by summarizing what is know about the state of basic science in the Eastern-bloc. We pointed to various limitations in existing sources of data, and proposed a somewhat different methodological framework for East-West comparisons. We then used this to assess the relative scientific performance of the principal Eastern-bloc high-energy physics accelerators. Even after making appropriate allowances for the likely bias in the various partial indicators employed, it was impossible to avoid the conclusion that the scientific outputs from each Eastern-bloc accelerator have been small in comparison with the nearest equivalent Western facilities. Of the main reasons for this discrepancy, some are technical in nature (inferior research facilities, scientific instruments and computers), but even these seem to be merely symptoms of a wider systemic and organizational malaise — both at the level of the research laboratory (in the limited resources available and the way these are managed), and of Eastern-bloc science in general.

• NOTES

The authors gratefully acknowledge the support of the British Economic and Social Research Council in carrying out this research (under grant number HR7448/2), and the help so freely given by large numbers of high-energy physicists in both East and West. Thanks are also due to a number of colleges at SPRU, and to Owen Lock, Martin Lockett, Linda Lubrano, Andy Pickering and Zhores Medvedev, for providing useful critical comment on earlier drafts of this paper. Finally, the authors would like to express their gratitude to David Edge and five referees for considerable help in improving the paper's presentation. However, the conclusions remain the responsibility of the authors alone.

1. No order of seniority implied (rotating first authorship).

2. See, for example, the NATO Science Committee report, *Research Management and Zero Growth* (Brussels: NATO, 1981).

3. The evaluation techniques are described in detail in B. R. Martin and J. Irvine, 'Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, Vol. 12 (1983), 61-90. The techniques have been applied to two other fields: see Martin and Irvine, 'An Evaluation of the Research Performance of Electron High-Energy Physics Accelerators', *Minerva*, Vol. 14 (1981), 408-32, and Irvine and Martin, 'Assessing Basic Research: The Case of the Isaac Newton Telescope', *Social Studies of Science*, Vol. 13 (1983), 49-86.

4. See the work cited in note 3.

5. The roles of science and technology in Soviet industrialization are comprehensively discussed in R. Lewis, *Science and Industrialization in the USSR: Industrial Research and Development, 1917-1940* (London: Macmillan, 1979).

6. N. I. Bukharin et al., *Science at the Crossroads* (London: Frank Cass, 2nd edn, 1971). The 1930s 'radical science movement' in Britain is discussed in detail in P. G. Werskey, *The Visible College* (London: Allen Lane, 1978). Interest in Soviet science was further stimulated by the publication of J. D. Bernal's book, *The Social Function of Science* (London: Routledge, 1939).

7. See, for example, *Radical Science Journal* and the British Communist Party magazine, *Science Bulletin*; also the British Society for Social Responsibility in Science publication, *Science for People*, and its American counterpart, *Science for the People*.

8. One of the first collaborative agreements was in the field of high-energy physics. It was drawn up in the late 1950s, and involved the European organization for Nuclear Research (CERN) at Geneva, and the East European Joint Institute for Nuclear Research (JINR) at Dubna in the Soviet Union. For details, see W. O. Lock, *A History of the Collaboration between the European Organization for Nuclear Research (CERN) and the Joint Institute for Nuclear Research (JINR) and with Soviet Research Institutes in the USSR 1955-1970* (Geneva: CERN, 1975). The history of US-USSR scientific co-operation is examined in L. L. Lubrano, 'National and International Politics in the US-USSR Scientific Co-operation', *Social Studies of Science*, Vol. 11 (1981), 451-80.

9. For example, T. Gustafson wrote a useful summary paper on the problems of Soviet Science, 'Why Doesn't Soviet Science Do Better?', in L. Lubrano and S. Solomon (eds), *The Social Context of Soviet Science* (Boulder, Col.: Westview Press, 1979), 31-67, as a result of participating in the US-Soviet Joint Working Group on Science Policy.

10. One of the best surveys of Soviet R&D personnel deployment is L. E. Nolting and M. Feshbach, 'R&D Employment in the USSR', *Science*, Vol. 207 (1 February 1980), 493-503.

11. A good example is S. Kassel, *The Relationship between Science and the Military in the USSR* (Washington, DC: Rand Corporation, R-1457-DDRE/ARPA, 1977).

12. For example, see the many studies referenced in C. P. Ailes and F. W. Rushing, *The Science Race* (New York: Crane Russak, 1982).

13. Two comprehensive and influential emigré critiques of East European science are provided by Z. A. Medvedev, *Soviet Science* (Oxford: Oxford University Press, 1979), and M. Popovsky, *Manipulated Science* (New York, Doubleday, 1979).

14. Reasons for the limited access to the normal mechanisms of informal scientific

communication include restrictions on travel, conference attendance, and the exchange of research papers (particularly early reports on 'work in progress'), the difficulty of making long-distance and international telephone calls, and so on. See, for example, Y. Levich, 'Trying to Keep in Touch', *Nature*, Vol. 263 (30 September 1976), 366-67.

15. Interlaced with emigré accounts are the perhaps inevitable occasional exaggerations and distortions. Indeed, these sometimes give rise to accusations of misinformation — see, for example, the review of Popovsky (op. cit. note 13) by C. Holden, 'Emigré Paints Grim Picture of Soviet Science', *Science*, Vol. 205 (7 September 1979), 981-84. Such accounts must clearly be used only with great caution.

16. See C. Kaysen (Chairman), *Review of US-USSR Interacademy Exchanges and Relations* (Washington, DC: National Academy of Sciences, 1977), and R. L. Garwin (Chairman), *Review of the US/USSR Agreement on Co-operation in the Fields of Science and Technology* (Washington, DC: National Assembly of Sciences, 1977). A useful summary is given in L. R. Graham, 'How Valuable are Scientific Exchanges with the Soviet Union?', *Science*, Vol. 202 (27 October 1978), 383-90. See also Lubrano, op. cit. note 8.

17. A study was also undertaken in which eminent American scientists familiar with Soviet work were asked to evaluate the health of their specialties in the USSR.

18. See Kaysen, op. cit. note 16, 100.

19. Graham, op. cit. note 16, 383.

20. See the first chapter, 'International Position of US Science and Technology', in National Science Board, *Science Indicators 1972* (Washington, DC: US GPO, 1973). These indicators have been extended in scope in subsequent volumes, and Computer Horizons Inc., who compiled the data under contract from the National Science Foundation, have produced comprehensive digests of the main indicator series on computer tapes.

21. F. Narin and M. P. Carpenter, 'The Adequacy of the Science Citation Index as an Indicator of International Scientific Activity', *Journal of the American Society for Information Science*, Vol. 32, No. 6 (November 1981), 430-39.

22. Gustafson, op. cit. note 9, 31. An interesting recent study of highly-cited papers is described in the research note by F. Narin, J. D. Frame and M. P. Carpenter, 'Highly Cited Soviet Papers: An Exploratory Investigation', *Social Studies of Science*, Vol. 13 (1983), 307-19.

23. To the extent that even the British Library Lending Division coverage of physics journals is unlikely to be 100% complete, these figures, too, are likely to be more biased against the Soviet Union than the US because of the easier access to journals in the latter.

24. Narin and Carpenter, op. cit. note 21. See also J. D. Frame and D. R. Prokrym, 'Counts of US and Soviet Science and Technology Journals', *Scientometrics*, Vol. 3 (1981), 159-75.

25. Narin and Carpenter, op. cit. note 21, 438-39.

26. See H. Schmied, 'A Study of Economic Utility from CERN Contracts', *IEEE Transactions on Engineering Management*, Vol. EM-24 (1977), 125-38.

27. This figure on the estimated world expenditure on high-energy physics comes from Table 4 in B. R. Martin and J. Irvine, 'CERN: Past Performance and Future Prospects — I. CERN's Position in World High-Energy Physics', *Research Policy*, Vol. 13 (1984), 183-210.

28. See the work cited in note 3 for details.

29. See footnote 8 in Martin and Irvine (*Minerva*), op. cit. note 3.

30. For a discussion of the patterns structuring the postwar distributions of basic research expenditures in the West, see J. Irvine and B. R. Martin, 'What Direction for Basic Scientific Research?', in M. Gibbons, P. Gummett and B. M. Udgaonkar (eds), *Science and Technology Policy in the 1980s and Beyond* (London: Longman, 1984), 67–98.

31. For example, S. Weinberg, a distinguished theoretical high-energy physicist, and future Nobel Prize winner, wrote in 1973, 'To me, the reason for spending so much effort and money on elementary particle research is not that particles are so interesting in themselves . . . but rather that as far as we can tell, it is in the area of elementary particles and fields (and perhaps also of cosmology) that we will find the ultimate laws of nature, the few simple general principles which determine why all of nature is the way it is': 'Where We Are Now', *Science*, Vol. 180 (20 April 1973), 276. See also various similar claims in L. C. L. Yuan (ed.), *Nature of Matter: Purposes of High Energy Physics* (Brookhaven National Laboratory, BNL-888, 1965): for example, H. A. Bethe, 'High-energy physics is undoubtedly today the frontier of physics. The discoveries in this field of study contribute most to the advance of our fundamental understanding of nature' (ibid., 9); and R. G. Sachs, 'High-energy physics is the principal operational arm of particle physics, which, in turn, is the essence of today's science of physics' (ibid., 20).

32. R. R. Wilson, 'The Next Generation of Particle Accelerators', *Scientific American*, Vol. 242 (January 1980), 39.

33. The only exceptions are three high-energy electron-positron colliders (PETRA at Hamburg, CESR at Cornell, and PEP at Stanford) which have been excluded because they came into operation only very recently, and because, for PETRA and PEP, there is as yet no Soviet equivalent.

34. JINR is the Eastern bloc equivalent of CERN, the European Organization for Nuclear Research at Geneva.

35. The *Zero Gradient Synchrotron* (*ZGS*) at Argonne was built as a direct competitor to the Dubna machine, with funds obtained during the 'cold war' period of US–Soviet competition in the 1950s. This is discussed in Comptroller General of the United States, *Report to the Congress: Increasing Costs, Competition May Hinder US Position of Leadership in High Energy Physics* (Washington, DC: US General Accounting Office, EMD-8058, 1980), 15.

36. An assessment of the scientific performance of these various electron accelerators can be found in Martin and Irvine (*Minerva*), op. cit. note 3.

37. See Comptroller General of the United States, op. cit. note 35. Some of the Western high-energy physicists whom we interviewed shared this conclusion, one putting it rather more succinctly: 'To a first approximation, Soviet accelerators do not exist'.

38. An extensive discussion of the technical and conceptual problems of using publication counts to evaluate scientific performance is given in Martin and Irvine (*Research Policy*), op. cit. note 3. The same paper also analyzes the problems involved in using citation analysis and systematic peer-evaluation data as indicators of relative scientific contributions.

39. See D. Sullivan, D. H. White and E. J. Barboni, 'The State of a Science: Indicators in the Specialty of Weak Interactions', *Social Studies of Science*, Vol. 7 (1980), 182.

40. These figures on the outputs from each accelerator cannot be obtained from the

BEST COPY AVAILABLE

ISI data-base, which lists the institutional affiliation of the authors of each paper but not the research facility where the work reported was carried out. This obviously greatly reduces the utility of ISI data for science-policy purposes.

41. These various approximate estimates were derived partly from an examination of the papers listed in international data compilations, and partly from an analysis of the experimental publications cited in Eastern-bloc papers.

42. This estimate was arrived at by examining a sample of papers to identify those with virtually the same title published in Russian and English.

43. Because the number of experiments that can be carried out on a collider at any one time is very limited compared with the corresponding number for a fixed-target accelerator, colliders tend to produce fewer publications.

44. Martin and Irvine (*Research Policy*), op. cit. note 3.

45. Note that even a paper subsequently found to contain 'mistaken' results can sometimes have a significant impact on the advance of knowledge if it stimulates research that might not otherwise have been carried out, and its citation record will normally reflect this.

46. This assumes that papers in journals not scanned by ISI contain roughly similar numbers of references to those in scanned journals.

47. This bias refers to the figures on total citations and numbers of highly cited papers. Of course, there would be no effect on the number of citations per paper, assuming that omitted publications were cited with the same average frequency as publications included in our list.

48. A further reason to assume that national and Soviet Republic journals are on average cited less is that, according to interviews carried out in Eastern Europe, researchers seem to prefer to place their better papers in the international journals when this is possible.

49. This is based on the assumption that excluded publications earn on average half as many citations per paper as those included in our list. This is probably erring on the generous side.

50. These data on the average number of references per paper for different journals came from the final volume of the 1980 edition of the *Science Citation Index* (Philadelphia: Institute for Scientific Information, 1980).

51. One other potential source of bias concerns the relatively frequent use by Eastern bloc authors of references to preprints rather than published articles. This may be partly because of the often long delays before papers are finally published, and partly because many Eastern-bloc groups do not have easy access to certain journals and instead have to rely on informally circulated preprints. Citations to preprints listed in the *SCI* have all been credited to the papers that supersede them, the same procedure being adopted for references to articles 'in press' and 'to be published'. By using such a procedure, it seems doubtful that any appreciable systematic bias against Eastern-bloc papers is introduced.

52. The citation rate provides an indicator of the average impact per paper from each accelerator, and thus allows for differences in the scale of research activity at each facility. Hence a small but relatively successful accelerator, even if it produces too few papers to gain a large total number of citations, may still be seen as making a significant contribution to science if it achieves a relatively high citation rate.

53. If anything, the omission of what are generally less important papers from the national and Soviet Republic journals will in all likelihood serve slightly to increase the average citation rate.

54. Consequently, there seems to be little evidence from these figures to support the view expressed in some interviews that Eastern-bloc researchers tend to publish longer, more substantial papers than their colleagues in the West (who might have been assumed to be more prone to the 'publish or perish' syndrome, and hence to splitting up results to publish them in several smaller, less substantial papers).

55. Justification for this can be found in Martin and Irvine, op. cit. note 27.

56. Conversely, papers published in national and regional journals tend to have a much narrower audience, and therefore stand far less chance of having a major impact and being highly cited. This is particularly true, for example, in the case of papers published in the physics journals of Soviet Republics.

57. This paper reported the first indications of a rising total cross-section in hadron-hadron interactions, a result later confirmed at the CERN ISR and one which contradicted the predictions of certain fashionable theoretical models of the time.

58. This does not include a number of other interviews that were terminated prematurely when it became apparent that the interviewee had insufficient knowledge to answer the questions satisfactorily, being unable, for example, to recollect correctly which accelerators were responsible for certain experimental results.

59. Where, for example, two accelerators were ranked first equal, they were each given the ranking 1.5; where three were placed first equal, they were ranked 2 (that is, the average of 1, 2, and 3), and so on.

60. The Kharkov and Novosibirsk machines were not included in this ranking exercise.

61. These lower-level contributions generally arise from experiments reporting more precise measurements of known particles and their properties without discovering anything new.

62. See note 43.

63. This is excluding the CERN SPS accelerator, which has only been operating since the mid-1970s.

64. Further support for this conclusion comes from the results of the attitude survey described in the concluding section. None of the East Europeans interviewed agreed with the statement, 'Serpukhov has been responsible for a significant number of major discoveries compared with other major high-energy physics centres', while 94% disagreed.

65. This was probably largely because of technical problems — see note 71.

66. A similar conclusion was also evident in the more general survey of opinion on Soviet high-energy physics presented in pages 100–02 of the Kaysen Report, op. cit. note 16. In discussing Novosibirsk, it is noted that 'although bright new ideas in particle technology (e.g. collective acceleration and electron cooling) have originated in the Soviet Union, the work has been of an undisciplined character, and the new initiatives have been more effectively followed up in the laboratories in the West than in those where they originated' (ibid., 100).

67. In the attitude survey, 76% of the East Europeans agreed to some extent with the statement, 'Soviet high-energy physics accelerators have been responsible for few major experimental discoveries', while only 18% disagreed.

68. See J. Irvine and B. R. Martin, 'CERN: Past Performance and Future Prospects — II. The Scientific Performance of the CERN Accelerators', *Research Policy*, Vol. 13 (1984), 247–84.

69. According to those interviewed, one of the strengths of Novosibirsk during the

1960s and 1970s was that its Director was generally able, by dint of his political and industrial contacts, to ensure that industry did supply the equipment needed. Other Soviet laboratories were not so fortunate.

70. See Schmied, op. cit. note 26.

71. Accelerators at the other Eastern-bloc laboratories have also experienced technical problems. Interviewees cited the Moscow proton synchrotron, and the Dubna synchrophasotron as being particularly afflicted by major technical limitations such as low beam-intensity and poor-quality external beams. Even at Novosibirsk, the Soviet centre widely regarded as being the most technically advanced, it was reported in 1972 that 'there had been difficulties in obtaining a good quality reliable beam on the VEPP 3 collider': *CERN Courier*, Vol. 12 (1972), 333. Three years later, it was noted that the performance of this machine still 'remains limited by inadequate positron intensities': ibid., Vol. 15 (1975), 73. This is partly a reflection of the fact that the main interest of the laboratory has been in accelerators rather than physics (both the first Director, Budker, and the present Director, Skrinsky, have made world-famous contributions to accelerator theory), and partly a consequence of the laboratory devoting a substantial fraction of its activity to the construction of accelerators for industrial and medical use — this has provided a means of helping to finance the series of VEPP machines.

72. There is, for instance, a fairly rigid distinction between, on the one hand, the Research Institutes of the Academy of Sciences and of the USSR State Committee for Atomic Energy where research is carried out but with limited staff, and, on the other, the universities, which concentrate largely on teaching. This is reflected in the Soviet participation in CERN experiments — almost all those involved are from the Research Institutes.

73. Of the East Europeans interviewed, only 18 percent agreed that 'Policy decisions relating to Soviet accelerators are determined strictly by scientific considerations rather than political ones', while 59 percent disagreed.

74. This is principally because the researchers interviewed generally had a far less intimate knowledge of the other Eastern-bloc accelerator laboratories, with the exception of Dubna, which has increasingly moved towards nuclear physics in recent years, and so is less of interest to the concerns of this paper than is Serpukhov.

75. Much of the impetus for this came from experimental work carried out at SLAC and the parallel theoretical endeavours of the American West Coast high-energy physics community. According to certain senior physicists interviewed, CERN was also perhaps a little slow to realize the significance of these developments. Given that the major contacts between Serpukhov and the West were focused around the formal agreement with CERN, this perhaps constitutes another factor underlying the delay before Eastern-bloc physicists began to appreciate the potential importance of the parton model.

76. For an interesting critique by a Western scientist of the way in which high-energy physics is currently administered and of the adverse effects of the intensively competitive pressures on younger researchers in particular, see various articles by R. J. Yaes: 'Physics Fads and Finance', *New Scientist*, Vol. 63 (22 August 1974), 462-63; 'New Generation Motivations', *Physics Today*, Vol. 29 (September 1976), 11; 'The Time Has Come to Abandon the Pumps and Run for the Lifeboats — Reflections on Leaving the Physics Profession to Study Medicine', *Bulletin of the American Physical Society*, Vol. 23 (1978), 21-22; 'The Science Establishment', in R. Arditti, P. Brennan and S. Cavrak (eds), *Science and Liberation* (Boston, Mass.: South End Press, 1980), 217-38.

77. When Western physicists were invited to submit proposals for experiments with the Mirabelle bubble-chamber at Serpukhov, these had to negotiate a rather tortuous path past not one committee but a sequence of four before final approval could be obtained: *CERN Courier*, Vol. 12 (1972), 209.

78. This example comes from one of our interviews. S. Weinberg and A. Salam in fact first proposed such a unified theory in 1967, but it was not until 1971, when G. t'Hooft showed the theory was renormalizable, that a large body of high-energy physicists began to take serious notice of it.

79. We do not wish to appear to be making a one-sided critique of the central institutions of Eastern-bloc society. There are, of course, a range of countervailing benefits to centralized planning. In particular, the socialist countries have at least been able to produce some semblance of rational mechanisms for determining priorities in science and technology, something which the Western nations are only now attempting to achieve. At the level of scientific practice, there is also less emphasis in the socialist countries on the ruthless competition that imposes tremendous burdens on many young researchers as well as distorting personal and social relations (see note 76). Nor is there much evidence of the rather wasteful competition between large laboratories to build equivalent research facilities first — as, for example, during the late 1970s, in the race between SLAC and DESY to build PEP and PETRA — a process which inevitably results in 'losers' and 'winners'. This is something which deserves analysis in a separate article.

80. In 1982, one of us was invited to lecture on the techniques of research evaluation in Bulgaria. Discussions with senior officials revealed the existence of major worries in several COMECON countries over the direction and efficiency of R&D activities.

John Irvine and *Ben Martin* are Fellows of the Science Policy Research Unit at the University of Sussex, where they work on a range of issues connected with policies for basic and applied research. They have in recent years acted as consultants to research-funding agencies in several countries and are the authors of a number of papers and reports on research policy, including *Foresight in Science: Picking the Winners* (London: Frances Pinter, 1984). *Authors' address:* Science Policy Research Unit, Mantell Building, University of Sussex, Falmer, Brighton, BN1 9RF, UK.

NATURE VOL.316 15 AUGUST 1985

Charting the decline in British science

from John Irvine, Ben Martin, Tim Peacock and Roy Turner

Recent data on the comparative scientific performance of different nations suggest that the decline in British basic science apparent during the 1970s began to level off at the start of the 1980s. However, rather than heralding the beginning of a revival, the evidence suggests that this is merely a temporary halt to the long-term slide.

IT is now almost conventional wisdom that British science is in a state of crisis. Recent leading articles in *Nature*, for example, have had a tone of desperation, with headlines such as "University paymasters discredited — the British government's latest policy document on higher education is a disgrace", "More crisis for UK research", and "Parliament's chance to act". A similar, if more restrained, note of despair has been sounded by the Advisory Board for the Research Councils (ABRC) in its latest report, *Science and Public Expenditure 1985*, which was submitted in June to the Secretary of State for Education and Science. In unambiguous terms this stated that, "We believe that Parliament, the Government and the country as a whole are complacent about the current financial straits of the dual support system" (for financing research in universities).

In these and other analyses of the situation now confronting British science, most attention has been focused on inadequacies of the funds and other inputs to research. Largely for lack of data, less has been said about the effects of financial constraints on the outputs from science. We attempted to remedy this last year in a study for the ABRC which made use of a range of research output indicators (world shares of publications and citations, and figures on Nobel Prizes and patents), seeking thus to provide systematic information on long-term trends in British science.

The evidence suggested that, during the latter half of the 1970s, Britain's basic science declined appreciably, both in regard to relative output of publications and in their impact on the international research community. It was argued that the government may not be providing sufficient support for research (or at least not concentrating it in the right places) to ensure the successful development of new core technologies vital to the long-term survival of British industry.

The subsequent public interest in the results of the study (see *New Scientist*, 1429, pp.25–29, November 1984) apparently reflects a growing realization that, despite all the limitations of scientific output indicators, it will be necessary to make much greater use of systematic information on research performance in formulating a coherent national research policy if Britain is to ensure that the scientific needs of new technology-based industries are met. Trying to steer publicly funded research using only the mechanism of scientific peer-review (or "the informed prejudices of wise men", as it has been aptly described), while totally ignoring quantitative input and output indicators, is somewhat like trying to run the economy using only the judgements of a committee of senior industrialists.

In what follows, we update the figures on Britain's performance in basic research to the end of 1982, and provide more detailed data on trends in some of the main fields of science. The work involved was supported by a consortium of British learned societies who have set up an informal working party (Ad Hoc Working Group for Research Funding, AGREF) under the chairmanship of Professor J. Lamb (St Andrews University, UK) to examine government support for science.

Measuring performance

Perhaps the most readily accessible data on national research outputs are the publication and citation indicators contained in the National Science Foundation's *Science Literature Indicators Data Base*. This is compiled by CHI Research on behalf of NSF for use in the biennial US *Science Indicators* report, and is based on the *Science Citation Index* produced by the Institute for Scientific Information (ISI). (This database is the source of the material presented in Tables 1–7, and a copy of it is held at the Science Policy Research Unit, University of Sussex.) The primary dataset covers some 2,000 of the world's leading scientific journals over the decade 1973–82, and contains figures on numbers of publications and citations for more than 190 countries and geographical regions, broken down into eight main scientific fields (for example, physics) and, in the case of the publication data, into approximately 100 subfields (for example, optics).

The use of such bibliometric data to indicate research output in basic science entails a number of assumptions. One is that most new scientific knowledge is eventually encapsulated in the form of research articles in learned journals. If this is the case, it follows that aggregate figures on the number of British publications in specified fields compared with those from other countries should give some indication of Britain's relative scientific output and, more importantly, of trends over time. A second assumption is that citation counts (that is, the number of times papers are subsequently referred to by other scientists) can be used to gauge the impact of research findings on the scientific community.

At the outset, we list three important caveats. First, the ISI data are appreciably biased in favour of English-language journals, which means that the scientific contributions of West Germany, France, Japan and, in particular, the Soviet Union are underestimated. Second, the CHI/NSF trend data are based on an essentially constant set of journals established in 1973; papers published in entirely new journals are not covered, with the result that newly emerging or fast growing areas of science may be under-represented. This disadvantage is, however, offset by the fact that more reliable analyses of trends over time can be undertaken. (Since 1981, CHI Research has in fact compiled an alternative data series based on the 3,000 journals scanned by ISI in that year, but

Table 1 UK world share of publications by field of research, 1973–82

Field	Percentage of total					Av. ann. % change	Av. ann. % change	Total % change
	1973	1975	1978	1980	1982	1973–80	1980–82	1973–82
Biology	9.5	10.7	10.8	10.2	9.9	+1.1%	−1.5%	+4%
Biomedical research	9.4	9.6	8.8	8.4	8.4	−1.5%	0.0%	−10%
Chemistry	7.8	8.3	7.0	6.8	6.9	−1.8%	+0.7%	−12%
Clinical medicine	10.4	10.4	9.5	9.3	9.4	−1.5%	+0.5%	−9%
Earth and space sciences	8.1	9.5	8.9	9.5	8.9	+2.5%	−3.2%	+10%
Engineering and technology	10.7	11.0	9.4	9.5	8.6	−1.6%	−4.7%	−20%
Mathematics	7.4	8.2	7.0	7.0	7.0	−0.8%	0.0%	−5%
Physics	7.7	7.4	6.3	5.9	6.1	−3.3%	+1.7%	−21%
All fields combined	**9.2**	**9.5**	**8.6**	**8.3**	**8.3**	**−1.4%**	**−0.1%**	**−10%**

their analysis suggests that this larger database may be more, rather than less, biased in favour of English-language countries.) Third, since CHI Research assigns articles to fields and subfields according to the journals in which they are published, some problems can arise in relation to smaller specialities, in particular those of an interdisciplinary nature where researchers publish in a great variety of journals.

Britain's performance

Table 1 contains data on changes over the decade up to 1982 in Britain's world share of publications broken down by field of research. While there has been a decline of 10 per cent in overall world share (down from 9·2 per cent to 8·3 per cent), the picture is uneven. Physics and engineering have suffered significant decreases of 21 per cent and 20 per cent respectively, but biology has marginally increased its world share by 4 per cent over the 10 years. More encouraging, at least at first sight, is that the downward trend evident since 1975 seems almost to have levelled off in the two years since 1980, with physics, chemistry and clinical medicine apparently registering small increases. We shall return to the question of whether this halt in the downward trend is likely to continue or is merely a temporary phenomenon.

On the impact of British work on the international research community, a rather less optimistic picture emerges from the citation figures in Table 2. It can be seen that the decline which began in the 1970s has continued into the 1980s — the world share of citations for all fields combined has decreased by 15 per cent since 1973, with biomedical research, engineering and physics all recording falls of over 20 per cent. (Figures based on the 1981 journal-set are relatively consistent with these values.) Furthermore, although

Table 2 UK world share of citations by field of research, 1973–82

Field	Percentage of total					Av. ann. % change	Av. ann. % change	Total % change
	1973	1975	1978	1980	1982	1973–80	1980–82	1973–82
Biology	11.9	12.5	12.8	12.0	11.4	+0.1%	−2.5%	−4%
Biomedical research	11.7	10.4	9.0	8.5	8.4	−0.6%	−3.9%	−28%
Chemistry	11.2	10.2	11.5	10.3	9.8	−1.1%	−2.4%	−12%
Clinical medicine	12.8	12.8	11.5	11.3	11.5	−1.7%	+0.9%	−11%
Earth and space sciences	9.4	9.9	8.9	9.3	8.9	−0.2%	−2.2%	−5%
Engineering and technology	11.5	12.1	10.9	9.9	9.0	−2.0%	−4.5%	−21%
Mathematics	8.7	9.4	7.3	7.8	7.4	−1.5%	−2.6%	−15%
Physics	7.6	7.8	6.1	6.7	5.9	−1.7%	−6.0%	−22%
All fields combined	**11.2**	**10.9**	**10.2**	**9.7**	**9.5**	**−1.9%**	**−1.2%**	**−15%**

Table 3 Percentage changes in world publication share of major industrial countries by field, 1973–82

Field	Canada	France	FRG	Japan	UK	USA	USSR	Rest of world
Biology	+6%	−30%	−10%	+24%	+4%	−5%	−7%	+8%
Biomedical research	−10%	−27%	+7%	+64%	−10%	+5%	−7%	−5%
Chemistry	−17%	−17%	+11%	+23%	−12%	−6%	−11%	+12%
Clinical medicine	+4%	−6%	−13%	+62%	−9%	−2%	−17%	+5%
Earth and space sciences	+5%	−4%	+46%	+4%	+10%	−8%	−9%	+12%
Engineering and technology	−15%	+31%	0%	+45%	−20%	−1%	−31%	+22%
Mathematics	−24%	0%	+15%	+52%	−5%	−19%	+74%	+31%
Physics	−19%	+7%	+28%	+38%	−21%	−9%	−16%	+20%
All fields combined	**−8%**	**−9%**	**+2%**	**+40%**	**−10%**	**−3%**	**−15%**	**+9%**

Table 4 Percentage changes in world citation share of major industrial countries by field, 1973–82

Field	Canada	France	FRG	Japan	UK	USA	USSR	Rest of world
Biology	+13%	−2%	+13%	+32%	−4%	−5%	−45%	+6%
Biomedical research	−9%	+10%	+48%	+83%	−28%	−4%	−6%	+3%
Chemistry	−14%	+3%	+4%	+59%	−12%	−9%	−13%	+11%
Clinical medicine	+2%	+17%	+3%	+84%	−11%	−4%	−27%	+6%
Earth and space sciences	−6%	+9%	+66%	+35%	−5%	−3%	−34%	+12%
Engineering and technology	−23%	+8%	−11%	+91%	−21%	+4%	−51%	+30%
Mathematics	−10%	+19%	+3%	+53%	−15%	−13%	+5%	+42%
Physics	−23%	+13%	+35%	+69%	−22%	−15%	−23%	+37%
All fields combined	**−7%**	**+8%**	**+14%**	**+65%**	**−15%**	**−5%**	**−27%**	**+11%**

BEST COPY AVAILABLE

there had been a slight decrease in the rate of decline from 1.9 per cent per annum in the 1970s to 1.2 per cent in the early 1980s, there is little sign in the citation figures of the levelling off seen in the world share of publications.

When the figures for Britain are set against those for the other main industrialized nations, the implications of the long-term decline in Britain's basic science become all too apparent. Table 3 gives a breakdown by field of percentage changes in national shares of world publications over the period 1973–82. Japan has increased its overall share by no less than 40 per cent and the rest of the world by 9 per cent; the main losers have been Britain (−10 per cent), Canada (−8 per cent), France (−9 per cent) and the Soviet Union (−15 per cent), although in the last two cases there are likely to be appreciable biases in the statistics since French and Soviet scientists may well be preferring to publish more in domestic journals not scanned by ISI.

Such an effect would also give rise to an element of bias (probably rather smaller) in the figures on changes in national shares of world citations given in Table 4. These suggest that Japanese researchers in all eight fields have markedly increased their impact on the international scientific community, with an overall citation share that has risen by 65 per cent during the 10 years to 1982. Similarly, West Germany and the rest of the world increased their overall shares by more than 10 per cent, with France also recording a significant increase (8 per cent). Britain and the Soviet Union have again been the main losers, with declines of 15 per cent and 27 per cent respectively. One interesting feature of the table is that Japan's largest gains in world citation share have tended to be in fields where Britain has declined most, biomedical research, physics and engineering in particular.

One finding from our previous analysis of earlier NSF/CHI data was that the weakest parts of British basic science are heavily concentrated in areas with strategic technological importance for the future. This finding is reinforced by the updated statistics reported in Table 5. This table lists those scientific specialities in which Britain had a world publication share of more than 11 per cent or less than 7 per cent in 1982. One can see that the areas of strength are mainly in medical and biological specialities, although organic and inorganic chemistry, and electrical engineering and electronics are also quite strong. In contrast, the weaker areas are generally in more applied scientific and engineering specialities such as metallurgy, optics, polymer research, chemical engineering and applied physics. What is particularly alarming is the continuing rapid decline in several of these areas; in just two years since 1980, Britain's world publication share in optics dropped from 5.5 per cent to 4.6 per cent, in polymer research from 6.3 to 5.5 per cent and in chemical engineering from 6.7 to 5.9 per cent.

More systematic evidence on recent trends is contained in Table 6, which shows the research areas where Britain has either increased or decreased its world publication share by over 15 per cent in the five years up to 1982. With the exception of computer technology, the overall picture is again one of major declines in areas of science likely to prove strategically important in the future — for example, materials science, applied chemistry and physics, and electrical engineering and electronics.

We have argued elsewhere that companies in both traditional and new industrial sectors are becoming increasingly dependent on the results of research for maintaining innovative activity and international competitiveness (see *Foresight in*

Table 5 Strengths and weaknesses in British science: share of world publications, 1982

Weaker areas: < 7 per cent of world share	% Share
Clinical medicine–pharmacy	1.2
General chemistry	3.7
General physics	4.3
Metallurgy	4.6
Nuclear and particle physics	4.6
Optics	4.6
Polymer research	5.5
Oceanography and limnology	5.6
General mathematics	5.7
Ophthalmology	5.7
Radiology and nuclear medicine	5.7
Chemical engineering	5.9
Nuclear technology	6.0
Solid-state physics	6.0
Applied physics	6.1
Agricultural and food science	6.4
Cardiovascular research	6.7
General biomedical research	6.8
Analytical chemistry	6.9
Surgery	6.9

Stronger areas: > 11 per cent of world share	% Share
Anaesthesiology	18.4
Pathology	15.6
Tropical medicine	15.1
Ecology	14.6
Veterinary medicine	13.3
General and internal medical research	13.0
Inorganic and nuclear chemistry	13.0
Probability and statistics	13.0
Gastroenterology	12.9
Dermatology and venereal disease	12.7
Electrical engineering and electronics	12.7
Hygiene and public health	12.5
Dentistry	12.2
Microbiology	12.0
Geology	11.9
Haematology	11.8
Botany	11.3
Organic chemistry	11.3

Table 6 Major recent trends in British share of world scientific publications, 1977–82

Increase of ≥ 15 per cent in world share, 1977–82	% Increase
Entomology	70
Computer technology	61
Dentistry	17
Gastroenterology	17
Geology	16
Immunology	15
Haematology	15

Decline of ≥ 15 per cent in world share	% Decrease
Metallurgy	34
Clinical medicine–pharmacy	32
Materials science	31
Chemical engineering	26
Tropical medicine	26
Nuclear and particle physics	26
Cell biology, cytology and histology	24
Polymer research	24
Veterinary medicine	23
General physics	22
Electrical engineering and electronics	21
Applied chemistry	19
Ecology	19
Applied physics	17
Pathology	17
Botany	16
Dermatology and venereal disease	16
Nutrition and dietetics	16
Endocrinology	15

BEST COPY AVAILABLE

Science: Picking the Winners, Pinter, London, 1984). The figures on recent trends in British science, especially those for more applied scientific and engineering research, thus give major cause for concern, particularly when set in the context of the dramatic fall that has already occurred over the last decade or so in the country's relative technological standing. One indication of the latter is the British share of foreign patents registered in the United States. OTAF data shows that, in the past 20 years, this has declined from around 20 per cent to 8 per cent, while Japan's share has, over the same period, risen from under 10 per cent to over 35 per cent, and West Germany's has remained broadly constant at about 25 per cent, as has that of France at 8 per cent. Clearly, if Britain is to compete more effectively in the international technology market, one important precondition will be an increase in the strength of its underlying research infrastructure.

Optimism

It is therefore worth looking more closely at what has happened since 1980 to determine whether there is any evidence that the long-term decline in British science shows any sign of being reversed. As noted earlier, the figures relating to the British world share of publications do initially seem to provide some grounds for optimism. From the figures in Table 1, one can see that, while the overall share fell from a peak of 9.5 per cent in 1975 to 8.3 per cent in 1980 (a fall of 14.5 per cent in five years or nearly 3 per cent per annum), the British share has remained approximately constant since 1980. Does this herald a renaissance in British science?

There are several possible explanations for this levelling-off in the relative output of British publications since 1980. One is that British scientists, perhaps imbued with a new sense of economic realism engendered by the policies of the Conservative government after its election in 1979, realized the need to increase their productivity and began to publish more research articles in international journals.

In our view, however, a more plausible explanation relates to the funds received by science. Most British academic research is supported by grants from the five research councils with indirect support provided by the University Grants Committee. Figures on research council expenditures in constant prices (calculated using the official EEC research and development price deflator) for 1970 onwards have been plotted graphically in Fig. 1, the expenditure for each year being expressed as a ratio of that in 1977. As can be seen, research council expenditure grew in real terms between 1970 and 1972, before declining gradually to a minimum in 1977. In the three years up to 1980, expenditure grew slowly but steadily, but since then has fallen sharply in real terms.

Figure 1 also plots the changes in Britain's world share of publications and citations. In the case of publications, there was a peak in 1975 (three years after the peak in funding), but then a rapid fall reaching a minimum in 1980 (again three years after the trough in funding); 1981 registered a modest increase, while in 1982 the British world share began to fall again. There is thus a certain similarity between the trends over time in funding and national share of publications. It would appear that the 6 per cent increase in funding between 1970 and 1972 was reflected in the 3 per cent increase in Britain's share of publications between 1973 and 1975, while the 7 per cent decline in funds in the five years up to 1977 was associated with the sharp drop of 14 per cent in publication share between 1975 and 1980 (see Table 7). In other words, there seems to be a time lag of approx-

Table 7 Association between research council expenditure and UK world share of publications

Research council expenditure		UK share of publications	
1970–72	+6%	1973–75	+3%
1972–77	−7%	1975–80	−14%
1977–80	+3%	1980–82*	0%
1980–83	−6%	1983–86	?

* 1983 figure not yet available

imately three years between changes in expenditure and publication share. (This is the reason why the figures on publications have been normalized to 1980 compared with 1977 for funding.) Similarly, there is some suggestion of a time lag between publication and citation world shares, with the rapid fall in the former between 1975 and 1979 apparently being mirrored in a similarly sharp decline in the latter between 1976 and 1980.

If there is indeed an association between research expenditure and the British share of world publications, this would suggest that the levelling off in publication output was largely the result of the 3 per cent increase in research expenditure between 1977 and 1980. Further, if such an association were to continue to hold in the future, one would expect the drop in funding since 1980 to be followed some three years later by a sharp decline in the UK share of publications. It will be interesting to see whether this is actually the case when the CHI/NSF data for 1983–84 become available in 1986. Furthermore, according to the ABRC's latest report, research council funds have continued to fall in real terms since 1983 (by perhaps 1 per cent per annum) and, on the basis of government estimates, will probably continue to do so over the rest of the decade. If the past is any guide, it seems that the UK world share of publications may decline at a somewhat faster rate over the foreseeable future.

Much more research is obviously needed to establish the existence of a direct relationship between funding and publication output and also to explore the reasons behind the fall in the British share of citations. (One reason, for example, may be the widely reported shortage of front-line research equipment in universities and research council laboratories.) Nevertheless, it is clear that one can take little comfort from what has happened since 1980 — the evidence suggests that the levelling-off in the decline of the UK world share of publications between 1980 and 1982 will prove only a very temporary phenomenon.

Over the last year, there has been a considerable growth in concern in Britain with the state of science. At the invitation of ABRC, the Royal Society is undertaking a study of the health of British science which may provide some evidence on the extent to which the pessimistic conclusions following from bibliometric data of the type presented above are consistent with the views of the scientific community. The Secretary of State for Education and Science is, according to his speech during the 14 June House of Commons debate on science policy, awaiting the completion of that study before making his judgement about whether the quality of science in Britain has declined. Others, however, are already extremely concerned. For example, John Harvey-Jones, chairman of ICI, had this to say about a thorough review by the company of where its next science-based innovations would come from: "I think there is some evidence that the United Kingdom is sliding in the world pecking order. The evidence is still tenuous, but it has led us to start increasing our links with overseas universities that we see as centres of excellence. . . . If our suspicions are correct that the United Kingdom is no longer up at the front as it was, then that presents, in the long haul, very dangerous problems for us." (Quoted in *The Guardian*, 19th February 1985, p.24.) As one of the Members of

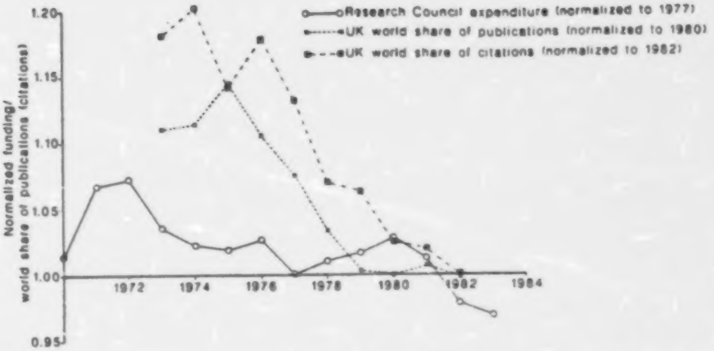


Fig. 1 Association between research council expenditure and UK world share of publications/citations. Source: Tables 1 and 2 on pp.6 and 19 of Butler, E. B. (ed.) *Research and Development in the UK* (Technical Change Centre, London, 1984), which are in turn based on figures published by the Central Statistical Office of the United Kingdom.

Parliament who spoke earlier in the parliamentary debate warned, "this may be the last chance . . . to stop the catastrophe that we face as a scientific and educational nation".

We acknowledge support from AGREF, the working group convened by a number of the learned societies, and from the Leverhulme Trust and the Economic and Social Research Council, as well as assistance provided by Dr F. Narin of CHI Research. ☐

John Irvine, Ben Martin and Tim Peacock are researchers at the Science Policy Research Unit, University of Sussex, Falmer, Brighton BN1 9RF. Roy Turner works in the Transport Studies Group at the Polytechnic of Central London.

The assessment of Government support for industrial research: lessons from a study of Norway

Michiel Schwarz, John Irvine, Ben Martin, Keith Pavitt, and Roy Rothwell
Science Policy Research Unit, Mantell Building, University of Sussex,
Falmer, Brighton, Sussex BN1 9RF, U.K.

Abstract: This paper summarises the results of a recent study carried out by the Science Policy Research Unit for the Norwegian Royal Commission on Industrial Research (the 'Industriforskningsutvalget'). We assess in broad terms the effectiveness of existing Norwegian mechanisms for supporting industrial research in institutes and firms, focussing in particular on the mechanical engineering and electronics sectors. A range of evaluation techniques are utilised in examining (a) research institute performance, (b) the research institute/industry interface, and (c) the general mode of operation of the funding system. It is suggested that the methodology adopted in the study may be of more general value in evaluating national systems for funding R & D.

INTRODUCTION

Governments are increasingly acknowledging both the importance of industrial innovation in relation to economic growth and competitiveness, and the necessity for government measures to encourage and support innovative activity (OECD, 1978). Concern about such questions has recently impelled several Western nations (for example, the Netherlands, Sweden, and the United States) to examine the appropriateness and effectiveness of existing structures and mechanisms for the state support of R & D (cf. Rothwell and Zegveld, 1981). Against a background of structural change in national economies and radical advances in technology, questions are being raised as to whether current structures of government support for industrial R & D are still operating effectively. This has in turn generated a need for appropriate methods of assessing R & D support systems, in order to determine whether significant improvements are required.

This paper reviews some of the methods, results and policy implications of a study on the Norwegian system of state support for industrial R & D, a study which involved a systematic assessment of certain major elements in that country's R & D system. The research was carried out over a 4½-month period in 1981 by a team from the Science Policy Research Unit (SPRU).

Though the resulting report (SPRU, 1981) is primarily concerned with one particular country, it may nevertheless be seen as illustrating many of the general problems associated with government support for industrial innovation, and with the assessment of R & D support mechanisms. The research focussed in particular upon policies for the creation and maintenance of an effective scientific and technical infrastructure, and for ensuring an appropriate distribution of R & D project funds to research institutes and industry. Given the requirement for new policy-analysis tools, the approach reported in this study may be of wider significance insofar as it can be applied in future evaluations of R & D activities and related government-industry links in other countries. Certainly, very few detailed empirical assessments of this sort have been carried out previously.

For reasons of space, it is not possible to reproduce here all the data reported in the study. Instead, we shall summarise the nature of these data, explaining the methods used to produce and analyse them, and discussing in detail those findings that are of general relevance to current problems in R & D management and research policy.

1. BACKGROUND AND FOCUS OF THE SPRU STUDY

The SPRU study was commissioned in November 1980 by the Industriforskningsutvalget, the Royal Commission set up by the Norwegian Department of Industry to identify the strengths and weaknesses of the present Norwegian industrial R & D system, and formulate proposals for improving it. The Commission reported in September 1981 (NOU, 1981).

An important focus of the Commission's work was the Royal Norwegian Council for Scientific and Industrial Research (NTNF). Of the various mechanisms available for the state support of industrial R & D, Norway has historically chosen to concentrate on the support of technical research institutes, combined with a government-funded system for industrial R & D projects carried out collectively by industry and the institutes. The rationale behind this was the need in post-war Norway to modernise the industrial base, much of which lacked expertise and capacity in R & D. Over the 35 years of its existence, NTNF has grown appreciably in size and complexity, establishing its own research institutes (there are currently 15, although some are more concerned with public sector than industrial R & D), and creating a large network of committees (some 27 at present) to determine and administer the distribution of R & D resources. In 1979, the overall NTNF budget was approximately 400 million kroner (US$ 87 million), representing about 20% of total government funding for R & D (approximately 2 billion kroner).

In common with most industrial countries, Norway has also initiated a range of other mechanisms for the support of industrial R & D – in particular the Industrial Fund and the Development Fund. Hence not only is the overall level of governmental support for R & D a particularly important issue, but so is the *balance* between these various mechanisms for the support of R & D. In particular, what proportion of R & D resources should be allocated to infrastructural support? To what extent is Norway's NTNF-based R & D system, with its emphasis on central institutes for collective technical research, able to meet the pressing needs of today's industry? This question is especially important given that Norwegian industry spends relatively little (as a proportion of GDP) on industrial R & D: in 1977, the Norwegian business enterprise sector spent the equivalent of only 0.49% of GDP on industrial R & D, less than half that in Japan (1.08%) and West Germany (1.12%), and significantly less than other small industrial nations like Switzerland (1.71%) and Sweden (1.10%).

A further reason for assessing the role and performance of NTNF stemmed from a concern expressed by certain critics of the system about the degree of complexity and inflexibility which, it is argued, has evolved over the years, and about the implications of this for the system's continued effectiveness. NTNF is both a foundation providing funds for research (in institutes and firms) and a network of institutes performing research. At a time when the national research budget is growing more slowly than in former years, is this system still able to respond flexibly to the future needs of industry?

Finally, worries have been expressed both within and outside the NTNF system concerning the extent to which funds are being used effectively, critics pointing, for example, to 'ageing' problems in research institutes and to a lack of accountability over the results of research. To what extent is such criticism valid? Are firms and institutes making the best possible use of increasingly scarce R & D resources?

In order to analyse and evaluate Norwegian mechanisms for the support of industrial R & D, the Royal Commission therefore needed a wide range of data on the performance of research institutes and the effectiveness of the system by which NTNF awards project grants to institutes and firms. They decided to commission the Science Policy Research Unit to carry out the task of providing these data and giving an 'outsider's' assessment of certain key elements of the R & D support system.

The SPRU research focussed in particular upon three areas:

(i) the performance of a number of individual research institutes and the relationship of their activities to industrial needs;
(ii) the system for allocating R & D grants to institutes and firms;
(iii) the overall performance of the NTNF system in supporting R & D and industrial innovation.

The assessment centred on four institutes and on their R & D activities within two industrial sectors: mechanical engineering and electronics. These institutes were the Central Institute for Industrial Research (SI), the Foundation of Scientific and Industrial Research at the Norwegian Technical University (SINTEF), Rogaland Research Institute (RF), and the Norwegian Defence Research Establishment (NDRE). The first three of these provide different 'models' for collective research with different organisat[illegible] and funding patterns, operational benefits and problems. SI is a large NTNF operated institute, while SINTEF, though dependent on NTNF grants, is autonomous and closely integrated with the Technical University at Trondheim; RF is a regional institute (linked to Rogaland Regional College) set up specifically to serve the needs of local industry. In addition, the Royal Commission asked that NDRE – which is concerned with providing technology primarily for a single user, the Ministry of Defence – be included for comparison with the three collective research institutes so that their relative strengths and weaknesses could be more clearly indentified. (The Royal Commission did not, however, request that any of the government-supported 'branch institutes' (which operate as collective research organisations in a number of specialised industrial sectors) be included in the evaluation.)

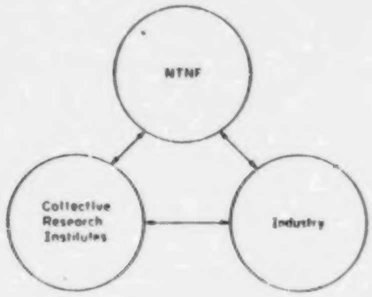


Figure 1 The NTNF-based research system

2 THE CENTRAL ROLE OF COLLECTIVE RESEARCH INSTITUTES

One significant characteristic of Norwegian research policy has been its emphasis on supply-side aspects of R & D, in particular (i) the scientific and technical infrastructure based on collective research institutes, and (ii) financial support for R & D projects in institutes and firms (cf. OECD, 1978). This, together with the fact that the flow of NTNF funds to firms takes place, to a considerable degree, through collaborative R & D projects carried out by research institutes on behalf of firms, means the effective operation of the overall R & D support system depends—perhaps more than in many other countries—on there being close links between institutes and firms. The SPRU study therefore focussed in particular on these links (Figure 1).

Research institutes have three main sources of funds:

(i) general or 'core' funding from NTNF;

(ii) grants from NTNF for individual projects, often for longer-term and more 'basic' research;

(iii) payment for R & D work commissioned by outside bodies, particularly government departments and firms.

The central role of collective research institutes is clearly illustrated by Table 1. In 1979, research institutes received 155.2 million kroner in general funding, the great bulk of the 127.1 million kroner awarded by NTNF for research projects undertaken outside industry, and 268 million kroner for projects sponsored principally by government departments and firms. In addition, a significant portion of the 46.7 million kroner awarded by NTNF in the form of project grants to industry was used by firms to pay for work carried out at institutes. Often, for example, a firm may be given a grant to help transfer to that firm the results of a long-term research project carried out by an institute. In

Table 1 Expenditure on NTNF-related research 1979

1. NTNF expenditure	million kroner
General funds to institutes	155.2 (38%)
Research grants—projects carried out in industry	46.7 (12%)
—other projects (carried out in institutes, universities, colleges, etc.)	127.1 (31%)
Other expenditure	74.8 (19%)
Total NTNF expenditure	403.8 (100%)
2. Other income to NTNF institutes (from government departments, industry, international collaborations, etc.)	268.0

Source: SPRU (1980) Table 4.

Note: It should be stressed that NTNF has over time devoted an increasing proportion of its funds to public sector research (on the environment, natural resources, regional development, and so on). The percentage devoted to *industrial* R & D has consequently fallen from about 90% in 1967 to 55% in 1979. Of this 55%, just over one fifth was allocated to firms.

Table 2 Critical issues relating to collective research systems

Critical issue	Main concerns
Funding	—general 'core' funding vs project-specific grants —collective research vs contract research for an individual sponsor —basic research vs strategic research —short-term funding vs long-term funding —state funds vs industrial funds
R & D Programme Selection	—short-term problem-solving vs long term open-ended research —modes of decision-making in selecting projects —identification of projects with greatest success potential —concentration of support on those firms in greatest need of R & D support
Personnel	—quality and quantity of recruits —mobility of staff (e.g. institute researchers moving temporarily to a firm to implement project results) —interchange of R & D personnel between universities, institutes and firms —development and maintenance of skills
Services	—the range of services offered by collective research institutes —availability of non-technical support for firms attempting to innovate —provision of assistance at different stages in the research, development and innovation process
Users	—identification of all potential users —specification of user requirements for R & D support —responsiveness to user needs, especially of small and medium-sized firms
Outputs and Implementation of Results	—information flows between collective research system and industry —industrial 'awareness' about range of support services offered —information on current and future needs of all potential users of results from collective research —transfer of research results to user firms and to industry generally —mechanisms for quality control of output

Source: based on Rothwell (1980)

such cases, NTNF generally makes it a formal condition of the grant that the firm contributes an equivalent amount to the work.

This comparatively heavy concentration of resources in research institutes gives rise to one of the principal sources of strain within the R & D support system—namely, the feeling among firms that they receive far too small a percentage of the total NTNF expenditure (under 12% in 1979) compared to the institutes. Not only is a large fraction (around 38%) of funds allocated directly to institutes which are formally part of the NTNF system and thus have established claims on the funds administered by NTNF; but in addition the other type of funds—project grants—is distributed to firms in direct competition with the institutes, since both submit project proposals to the same committees. Given the expertise accumulated within institutes and their close links with NTNF and its committees, many firms (and especially smaller ones) feel that they are at an unfair disadvantage in this competition, and that this is reflected in the resultant distribution of resources. For example, only 39% of project funds allocated in 1979 by the Mechanical Engineering Committee went to industry, while for the Electronics Committee the figure was only 29%. Moreover, the actual sums involved (7.2 and 4.7 million kroner respectively) were in absolute terms extremely small compared with the resources spent on R & D by firms in these two sectors—smaller indeed than the R & D budgets of many medium-sized firms. Also notable is the fact that large firms (over 500 employees) received over 60% of the grants awarded to industry.

Given the concentration of R & D funds on collective research institutes, it is worth noting some of the major characteristics of collective research as a mechanism for implementing innovation policy, particularly the problems which have been associated with its operation in other countries. Rothwell (1980) has summarised the critical issues here: they are listed in Table 2. Evaluation of the extent to which such problems existed within the Norwegian collective research system provided a focus for much of our data collection and analysis.

3. DATA COLLECTION AND ANALYSIS

In view of the central position of collective research institutes in the Norwegian R & D system, our prime task was to assess the effectiveness and appropriateness to industrial needs of the services offered by institutes. This required detailed information on the inputs (funding, staffing, etc.), outputs (long-term research, commissioned R & D work, support for industry in implementing new technology etc.), and overall mode of operation of the institutes, as well as data on the specific R & D requirements of industry. There data were generated largely through structured interviews with institute researchers and industrialists, with additional information provided by institute management, NTNF administrators, and government officials.

The main types of data used in our assessment are summarised in Table 3, these having

Table 3 Data collected on institute performance—a summary

Funding	R & D Programme Selection	Personnel
source and level of income	decision mechanisms used —longer-term projects —industrial projects analysis of main projects —initiator —nature of funding —scope —nature of work views on performance and selection —institute management and researchers —industrial managers	sources of recruitment qualifications of staff industrial experience of institute researchers links with industry degree of satisfaction felt by staff towards institute operation—'morale'
Services	**Users**	**Output/Results**
nature and extent of assistance offered views on adequacy of services —institute management and researchers —industrial managers characteristics of major projects views on state of technology and skill levels at institutes —institute management and researchers —industrial managers	nature of usage of institutes —product development —process development —other R & D functions usage by function and firm size % of firms not using institutes usage by function and industrial sector usage by function and firm's R & D intensity views of firms on utility of institutes views of institute researchers on institute utility	number of publications and citations for each institute views of firms on 'relevance' of longer-term research programmes at institutes views of firms on success of sponsored R & D projects variation in success rate of projects between institutes and for different types of project success rate factors —type of project —initiator (firm or institute) —size of firm —R & D intensity of firm variation in success rate factors between institutes variation in success rate factors between industrial sectors characteristics of least satisfactory projects as perceived by firms —for different sectors —for different project functions views of firms on improvements needed at institutes (perceived) overall strengths and weaknesses of institutes

Table 4 Information base for institute assessment: interviews conducted and R & D projects examined

1. Institute-based	SI	SINTEF	RF	NORE	Total
No. of staff interviewed	27	30	13	13	83
No. of R & D projects examined —long-term research projects	8	17	—	5	30
—shorter-term industrial R & D projects (at least partly funded by industry)	20	35	16	—	71

2. Firm-based	Electronics firms		Mechanical engineering firms			Other industrial interviews	Total
No. of managers interviewed	33		45			16	94
No. of industrial R & D projects examined	SI 30	SINTEF 29	SI 16	SINTEF 26	RF 12	—	113

been classified under the six types of critical issues relating to collective research identified in Table 2. Table 4 gives details of interviews carried out and R & D projects examined during the two months of field-work. In total, 177 interviews were undertaken (typically lasting 1½ to 2 hours), about half of which were with institute researchers and half with senior industrial managers. A further 11 interviews were undertaken during the initial exploratory phase of the study, enabling us to identify the key issues on which to focus, and providing a basis for drawing up the structured interview schedules used in later field-work.

At the institutes, we focussed on 30 long-term research projects. These consisted of the more 'basic' type of research projects, funded initially by NTNF grants, with public agencies and industry generally involving themselves at a later stage as the work progressed to the 'development' phase and the results were ready to be implemented in the form of new products or processes. We also examined 71 'industrial R & D' projects—that is, projects with far less research content, and funded from the outset by a public agency or a firm (sometimes with the aid of an NTNF grant).

In addition, both institute management and

Table 5 Data collected relating to R & D support and to the assessment of the overall NTNF system

	NTNF R & D Projects	NTNF R & D Support System
Finance	• distribution of NTNF committees' R & D project grants —to industry —to institutes, universities, and colleges • views on the present distribution of funds and on required changes —institute researchers —industrial staff • distribution of NTNF project grants and of government Innovation Plan grants —between different industrial sectors —between firms of different sizes	• main sources of NTNF income • expenditure by NTNF —general funds for institutes —research projects in institutes —research projects in industry • views on level of funding and distribution of funds in NTNF R & D support system —institute management and researchers —industrial managers.
Performance	• analysis of decision-making mechanisms • degree of satisfaction with R & D support given by institutes to industry —views of industrial staff • variation in success rates for R & D projects —by industrial sector —by firm size —by R & D intensity of firm	• views of institute management and researchers on improvements needed • views of industrial managers on major R & D support requirements of industry • variation in degree of satisfaction of industrial managers with NTNF system —by industrial sector —by firm size —by R & D intensity of firms

researchers, and industrial managers were invited to make assessments of the system for allocating NTNF project grants, and the appropriateness of the current state support system for R & D in Norway. These results were complemented by the data on NTNF funding, in particular analyses of the grants allocated by the Committees for Mechanical Engineering and Electronics. Table 5 summarises the nature of these data.

Table 6 Factors affecting the success of industrial R & D projects carried out by institutes: electronics and electrical engineering firms

	Institute A		Institute B	
	No of projects	% completed satisfactorily	No of projects	% completed satisfactorily
Type of project				
—Product development	22	50%	24	71%
—Process development	11	9%	10	50%
Initiator				
—Institute	15	53%	9	44%
—Firm	15	20%	20	75%
Size of firm				
<200 employees	13	38%	10	70%
>200 employees	17	35%	9	63%
R & D intensity[1] of firm				
—High	10	50%	8	83%
—Medium	14	36%	16	69%
—Low	6	17%	5	60%
Total	30	37%	29	66%

Source: SPRU (1981), Table 25.

[1] The R & D intensity of a firm is defined as the ratio of its expenditure on R & D in a given year to its sales in that same year. For electronics firms, the three categories of R & D intensity were: 'low' = 0 to 5%; 'medium' = 5 to 10%; 'high' = over 10%. For mechanical engineering firms, where R & D intensity is generally much lower, the corresponding figures were 0 to 2%, 2 to 5% and over 5%.

Explanatory note

As can be seen, Institute A exhibits a relatively low satisfaction rate for process-development projects in particular, and for projects initiated by firms and projects for firms with a low R & D intensity. On the other hand, this institute has been more successful with product development, with projects that it, rather than firms, has initiated, and with work for high R & D intensity firms. The apparent reason for this is that Institute A has concentrated its electronics research efforts on the development of certain microelectronics products, using the funds and competence built up during a major programme funded over the last 13 years by NTNF. Initiated by the institute, this programme has tended to involve only a small number of firms, particularly those employing former researchers from the institute. Indeed, most of the efforts have been concentrated on assisting just one firm which the institute hoped could be developed into a major Norwegian micro-electronics enterprise. Because of this concentration of effort, industrial projects in other areas of electronics, particularly for low R & D intensity firms, appear to have received proportionately less attention, and have therefore tended to be less satisfactory as far as the firms are concerned.

At Institute B, process development projects have also been completed less satisfactorily but, unlike Institute A, projects initiated by firms have been significantly more successful than those initiated by the institute. Overall, because Institute B has not concentrated its research efforts so much on one programme, it seems to have been able to allocate its resources more flexibly than Institute A, in particular to meet the challenge of projects initiated by industry. The fact that electronics projects which the institute itself has initiated seem to have proved relatively less successful provides some evidence to support the assertion by certain firms that its research staff, unless closely supervised, tend to adopt too 'academic' an approach. Such an approach may result in research publications contributing to the advance of scientific knowledge but it is less conducive to the creation of technology suitable for commercial implementation in new products or processes. (Although we have tended to focus on identifying the problem areas, it should nevertheless be noted that, in the view of most firms, Institute B has proved relatively successful in terms of providing R & D support for industry.)

Table 7 Improvements required at one institute (A) as judged by electronics firms

Number of projects assessed	30
Areas of improvement	% of firms suggesting each improvement
Better R & D staff	37%
Better project management	20%
Less academic approach	20%
Better contact with industry	20%
Keep to budgets	13%
More long term research specialisation	13%
Keep to schedules	7%
More temporary transfer of staff to implement projects in industry	7%
Staff with more industrial experience	7%
Better marketing/PR of institute	3%
Other improvements	13%

Source: SPRU (1981), Table 30

4. RESULTS: SOME ILLUSTRATIVE FINDINGS

Using these different sets of data, we were able to construct various performance indicators for both the collective research institutes and the overall R & D support system. Then, by contrasting the respective views of researchers and industrialists, it was possible to identify where improvements were felt to be needed, both at the level of R & D management and governmental policies.

4.1 *Institute Assessment*

The analysis of institute performance drew heavily on a critical review of their longer-term research activity. Particularly important here was an evaluation of the impact of these programmes, especially in terms of the transfer of knowledge and techniques to industry. In addition, examining the shorter-term R & D projects carried out at institutes enabled us to identify the types of firms making use of institutes, the nature of the R & D support they received, and the industrial benefits as perceived both by industry and institutes. (We were also able to establish which types of firms made little or no use of the facilities and services offered by institutes.) Certain conclusions could then be drawn about which projects are most likely to be completed satisfactorily by institutes, and hence about the improvements most needed. Table 6 presents a selection of some of these data, showing how the success rate for industrial R & D projects varies with the type of project, the initiator of the research, and the characteristics of the firm involved. In addition, from the industrial interviews, we obtained an indication of the elements perceived by firms as most problematic in relation to the performance of institutes. Table 7 gives an example of the results obtained here.

In order to go on to draw more general conclusions at the level of R & D management and policy, it is important to extend the analysis to include the determinants of success and failure in joint industry-institute projects, in particular institutional and organizational factors. These are discussed below in the context of the critical issues associated with collective research identified earlier (see Table 2).

(i) *Funding and programme selection.* Two main problems exist here. First, the dual role of NTNF as both a funding agency and a performer of R & D has given rise to certain areas of potential conflict between institutes and firms, especially where they are competing directly for the same funds. In the case of electronics, for example, the long-term concentration of resources on the micro-electronics programme carried out by Institute A in conjunction with one particular firm (see the explanatory note to Table 6), has inevitably limited the support available for the development of this crucial area of new technology in other institutes (and firms). According to those interviewed, this happened because the institute was first in the field, and there were then pressures to continue funding its research staff. This problem of flexibility in funding programmes is compounded by the lack of scope for outside determination of the way in which institutes use their core funds. More generally, many firms, because their project proposals must compete with those from institutes, feel that they are at a major disadvantage:

institutes are generally more skilled in writing strong proposals and are in a better position to know what sorts of proposals are most likely to be funded. Consequently, an application from a firm may be turned down, not because the proposed work is unpromising, but because the proposal is poorly written or argued. Small firms, and firms with only a limited in-house research capability, are at a particular disadvantage in this respect. Conversely, larger firms (particularly those integrated, formally or informally, into the NTNF committee structure) and those able to propose more 'interesting' research work are much more strongly placed. Secondly, partly because of the pressures within Norway towards the democratization and decentralization of decision-making, a large number of NTNF committees now exist to determine the distribution of resources, each with a widely-based membership. As a result, there has been a tendency for NTNF to give a small 'piece of the cake' to everyone, rather than make the hard strategic choices (at the national level) as to which areas of R & D and which institutes should be funded. For most areas of R & D, there is certain 'critical mass' or 'noise level'; if the resources available are less than this, little progress is made, and the return on research investment is likely to be limited. Our analysis of both funding patterns and interview data indicated that such a fragmentation of research efforts had indeed taken place in certain key areas of Norwegian R & D.

(ii) *Personnel.* Successful research depends ultimately on the skills and creativity of the researchers. Collective research institutes, if they are to remain productive on a long-term basis, need to have a continuous inflow of new people and ideas, and a constant development and evolution in the skills of their incumbent staff. SINTEF achieves this through its integration with the Technical University at Trondheim (its staff find the involvement with graduate students particularly useful in their research), and NDRE through military service (young scientists and engineers often prefer to spend their national service carrying out research rather than undergoing military training, and NDRE is able to recruit the best of these). SI, however, has no such immediate source of people and ideas, and this may well be the main factor underlying what was found to be a rather less successful R & D record than SINTEF's. In addition, NDRE pursues a deliberate policy of sending many of its staff overseas for further training in universities and high-technology firms. For a small country like Norway, which conducts less than 1% of all the world's R & D, substantial efforts must obviously be devoted to monitoring technological developments overseas, and diffusing the results to potential users in Norway. NDRE has found that by far the most effective way of doing this is not through monitoring the technical literature (though this is important), but by sending staff overseas for limited periods. NDRE's experiences suggest that the collective research system would benefit greatly from wider use of this approach.

(iii) *Services offered.* The institutes appear to have been rather less successful in meeting the needs of firms for process developments than they have with product developments. In addition, while they are often able to help in the early stages of the innovation process (carrying out background research), the support they provide at the later stages—for example, in the translation of technical advances into viable production prototypes—is rather more limited. Other limitations include the generally short-term nature of the assistance they provide, the absence of special help for small firms, and a neglect of the non-technical needs of firms, such as assistance with market-forecasting (see also Rothwell and Zegveld, 1982).

(iv) *Users.* The contact between firms, particularly smaller ones, and research institutes is limited by a number of factors. On the part of firms, there is a general lack of awareness about the nature of the services available, an uncertainty as to how to make contact with institutes and with whom, and problems with the complex process involved in applying for NTNF project funds or technical assistance. As for the institutes, many of their staff tend to be somewhat limited in their knowledge of the technical and commercial needs of firms. This results partly from a general lack of industrial experience amongst industrial researchers (with certain notable exceptions), and partly from their relatively infrequent contact with firms. A significant conclusion from our assessment was that such problems in dealing with small and medium sized firms (with less than 200 employees) were much less apparent in the case of RF, the smaller regionally-based research institute.

(v) *Output*. Even when institutes apparently do carry out R & D projects satisfactorily, the results are sometimes not commercially implemented by industry. Often, especially in the case of smaller firms, lack of subsequent commercial success is primarily due to the absence of in-house staff with the necessary expertise to take over the technology, and transform it into a new or improved product or process. In other instances, responsibility lies to a greater extent with institute researchers who, lacking industrial experience, are unable to produce research results suitable for commercial implementation; for example, they create a product that is technically over-sophisticated. The experiences at NDRE in particular suggest that successful transfer of technology is most likely to be achieved when firms are involved from the outset – i.e. from the initial research stage – and when institutes continue to provide assistance through the later stages of the development process. The best way of ensuring this is through the temporary transfer of staff between firms and institutes. Thus what might be described as 'the first law of technology transfer' is: 'the best mechanism for transferring technology is through the movement of people'.

4.2 *Assessment of the NTNF-based R & D Support System*

The interviews with industrial managers and institute researchers provided the main data used in assessing the operation and limitations of the existing R & D support mechanisms, particularly those centred around NTNF. They also formed the basis of our identification of the main improvements required. In particular, we were able to produce data on the extent to which the collective research system was perceived by firms to be operating successfully, and to relate the varying degrees of satisfaction to industrial sector, firm size and firm R & D intensity. Table 8 provides an example of how these data were analysed, and the sorts of results that were obtained. It shows that, in the mechanical engineering sector, small firms and firms with a low R & D intensity are far from satisfied with the NTNF system, and indeed that many of them had little or no knowledge of the system and the support it provides. (Similar data were produced for the electrical sector, and on firms' perceptions of the industrial support provided by the collective research institutes.)

These perceptions of industrialists could then be related to structural problems in the NTNF system apparent from data on patterns of resource distribution. They could also be related to the views of the institute researchers on the R & D support system. Again, we can briefly illustrate the approach adopted with an example: the perceived bias of the collective research system against smaller firms. Industrialists and institute researchers both believed the NTNF system to be really helping only the large firms. (57% of institute researchers advanced this view compared with 22% who disagreed.) Conversely, few (8% of researchers) saw the research institutes as successful in meeting the needs of smaller firms. The importance of firm size in determining the level of industrial assistance given by the R & D support system is reflected in the apparent bias towards larger firms in NTNF resource allocation: firms with over 200 employees received 83% of NTNF research grants awarded over the last 4 years in the mechanical engineering sector, and 67% of the grants in the electronics and electrical engineering sector. This is despite the fact that small firms account for a significant part of both these industrial sectors. The conclusion is that small

Table 8 Overall views of mechanical engineering firms on the NTNF system – effects of firm size and R & D intensity

Characteristics of firms / Overall Views on NTNF system	Size		R & D intensity		
	Smaller firms (< 200 employees) n = 20	Larger firms (> 200 employees) n = 22	Low (< 2% of sales) n = 18	Medium (2%–5% of sales) n = 17	High (> 5% of sales) n = 7
Good (no change needed) or satisfactory (but could be improved)	0%	55%	22%	29%	43%
Not satisfactory or highly unsatisfactory	60%	41%	56%	47%	43%
No views or no knowledge of system	40%	4%	22%	24%	14%

Source: SPRU (1981), Table 39b

Table 9 Convergence between the views of institute researchers and industrial managers

	Views of institute researchers	Views of industrial managers
% of industrial R & D projects completed satisfactorily by:		
(1) Institute A	59%	50%
(2) Institute B electronics activity	63%	66%
(3) Institute B mechanical engineering activity	85%	88%
% of interviewees believing that institutes are serving the R & D needs of industry:		
(1) well	27%	12%
(2) satisfactorily, but could be improved	38%	46%
(3) unsatisfactorily	35%	43%
% of interviewees believing that the NTNF support system is working:		
(1) well	1%	3%
(2) satisfactorily, but could be improved	25%	35%
(3) unsatisfactorily	74%	62%

Source: various tables and appendix 14 in SPRU (1981)

firms apparently cannot compete for support on equal terms with their larger competitors.

Finally, it is worth stressing that perhaps the main strength of the results on the performance of collective research institutes and the overall R & D system was the degree of convergence found between different elements of our analytical data. The notion that different 'partial indicators' of performance should yield broadly similar results was a key element of our assessment methodology. In this respect, Table 9 demonstrates how researchers and industrialists held very similar views on the extent to which industrial R & D projects have been completed satisfactorily by the various institutes. These two groups also shared similar perceptions of overall performance of the NTNF support system. This was somewhat less true of their views on the degree to which collective research institutes are succeeding in meeting the R & D needs of firms; institute researchers were inevitably rather more positive in their views on this than the firms, although even here the differences were not great.

4.3 *Possible Improvements to the Norwegian R & D Support System*

As a result of our analysis, we were able to identify several major areas where improvements are perhaps required in the Norwegian R & D support system. The major conclusions emerging from our assessment are as follows:

(i) *Organisational changes* should be made in the system by which NTNF operates both as a funding agency and performer of research, in order to reduce the level of direct institute-industry competition for resources and ensure greater flexibility in restructuring the activities of institutes when necessary.

(ii) *Greater concentration* of research efforts is needed to avoid fragmentation and to permit the formulation of longer-term strategic research programmes linked to national R & D needs.

(iii) The *technology-transfer function* of research institutes should be strengthened particularly through encouraging greater interchange of R & D personnel, thereby creating closer relations with other research centres and better links between firms and institutes throughout the various stages of R & D projects.

(iv) More special help for *smaller firms* should be provided in the way of funds and services available from the collective research institutes and NTNF.

(v) Greater attention needs to be paid to *likely industrial benefit* as a selection criterion for R & D project grants, rather than to intrinsic technological interest. Similarly, there should be greater emphasis upon R & D *quality control* and *project evaluation* to ensure that resources, once allocated, are utilized as effectively as possible.

(vi) Increased attention should be given to the provision of *non-technical assistance* (e.g. with market-forecasting, mangement of the innovation process, marketing, etc.) to firms attempting to convert R & D activity into successful innovations. This is especially important in the case of smaller firms.

(vii) Greater *cohesion* is required in *government policy* towards industrial R & D support. This would involve strengthening the links between the NTNF system and other mechanisms for encouraging industrial innovation, and ensuring that they are integrated with overall national industrial needs through more explicit innovation and industrial policies.

It should be stressed, however, that, while our research focussed upon the scope for enhancing the effectiveness of the Norwegian system for the support of industrial R & D, the present system is by no means a poor one. In particular, compared with many other countries, it is relatively successful in bringing together industry and research institutes, providing mechanisms both for ensuring the 'relevance' of strategic long term research undertaken by institutes and for transferring to industry the basic knowledge and techniques developed as a result of this work. Hence there already exists a solid base on which to build a more effective infrastructure for the support of Norwegian industry in the future.

5. THE WIDER SIGNIFICANCE OF OUR STUDY

Given the growing importance of effective national research, development and innovation policies, there is an urgent need to devise improved methods for assessing the performance of research groups (and institutes) and the effectiveness of governmental R & D support mechanisms. For this reason, the study reported here, and the methodology on which it is based, may be of more general interest to R & D managers and policy-makers.

Our assessment concentrated both on the inputs and outputs of the Norwegian (NTNF-based) R & D system, and on the operational structures and practices which together determine the system's overall effectiveness. At the centre of the methodology used in the study is the notion of converging performance indicators based on different methods of assessment: it is the convergence reflected in the views of different interest groups associated with Norwegian industrial R & D, and the consistency with the other statistical data, which suggest that some degree of reliability can be placed on our findings.

This methodology, it should be noted, draws on a long-term programme of work on the evaluation of research performance undertaken at SPRU by two of the authors (John Irvine and Ben Martin). The aim has been to develop policy analysis tools useful in evaluating past research performance, and has combined traditional input-output approaches with a range of social science techniques (especially structured interviews and attitude surveys.) Though first developed for the assessment of certain basic science specialties (for example, radio astronomy: see Martin and Irvine, 1982), this Norwegian study suggests that the approach can be successfully extended to evaluate areas of applied research and engineering, yielding results of potential relevance to national policy-makers and those concerned with R & D management. While the approach is rather labour-intensive, the cost is still small compared with national R & D budgets. (The all inclusive cost of the Norwegian study was approximately £22,000, i.e. less than 0.05% of the current NTNF annual budget.) To this extent, we believe that the work reported here does constitute a significant first step towards the development of research assessment techniques, and hence towards the eventual establishment of more effective national R & D policies.

REFERENCES

Martin, B. R. and J. Irvine. 'Assessing basic research: some partial indicators of scientific progress in radio astronomy', *Research Policy* (in press).

NOU (1981). *Forskning, Teknisk Utvikling og Industriel Innovasjon.*, NOU 1981: 30A,B, Oslo: Norges Offentlige Utredninger.

OECD (1978). *Policies for the Stimulation of Industrial Innovation*, Paris: Organisation for Economic Co-operation and Development.

Rothwell, R. (1980). *'Trends in Collective Industrial Research', Report to the Six Countries on Innovation*, Delft, the Netherlands: Six Countries Programme Secretariat (TNO, P.O. Box 215).

Rothwell, R. and W. Zegveld (1981). *Industrial Innovation and Public Policy*, London: Frances Pinter (Publishers) Ltd.

Rothwell, R. and W. Zegveld (1982). *Innovation and the Small and Medium Sized Firm*, London: Frances Pinter (Publishers) Ltd.

SPRU (1981). Irvine, J., B. R. Martin and M. Schwarz, with K. Pavitt, R. Rothwell. *Government Support for Industrial Research in Norway: A SPRU Report*, in NOU (1981) 30B, above.

ACKNOWLEDGEMENTS

The main body of the work reported here was carried out by John Irvine and Ben Martin, with Michiel Schwarz being involved during the final 1½ months. In addition, Keith Pavitt and Roy Rothwell acted as consultants to the project. The study was financed by the Norwegian Government Industriforskningsutvalget, with funds for further analysis and writing this paper being provided by the Leverhulme Trust. The authors are grateful to Hans Skoie, Johannes Moe (Director of SINTEF), and Kjell Roderburg (Director of SI) for providing detailed comments on an earlier draft of this paper. However, the conclusions presented here are the responsibility of the authors alone.

Note: Copies of the SPRU Report can be obtained (for £5.00 to cover postage and handling) from John Irvine (Science Policy Research Unit, University of Sussex, Brighton BN1 9RF, United Kingdom), to whom other correspondence concerning this paper should also be addressed.

APPENDIX 2

CRITIQUES OF IRVINE AND MARTIN'S METHODOLOGY AND A REPLY BY MARTIN AND IRVINE

NATURE VOL 311 6 SEPTEMBER [illegible]

European high-energy physics

Poor marks for enterprise

IN the thirty years since the establishment of CERN, the European Organisation for Nuclear Research near Geneva, the laboratory has produced only two "crucial discoveries" compared with a dozen or more at United States laboratories, whose total cost is similar, and may even be less.

This is the bottom line of a lengthy balance sheet, with many other conclusions and observations, drawn up in nearly two years' work by social scientists Ben Martin and John Irvine, of the Science Policy Research Unit at the University of Sussex, and published today (6 September). The two-part report is certain to be influential in this particular reappraisal of its membership of CERN.

The first CERN discovery categorized as "crucial" was in 1973 ("neutral currents", a new form of weak interaction). By then, US laboratories had discovered a second neutrino, the omega-minus (the beginning of the quark model), charge-parity-violating interactions, deep inelastic scattering (which indicated that the proton really had components, now known to be quarks) and more. CERN, on the whole, has been thorough but dull, say Martin and Irvine — CERN lacked "bzazz". But the second of CERN's crucial discoveries is very recent (the intermediate vector bosons *W* and *Z* found last year), is attributable to a new adventurousness at CERN and, according to the review, "to a certain extent the balance of power in experimental high-energy physics between North America and Western Europe seems to have reached a turning point".

Martin and Irvine are famed (or notorious) for their previous critical assessments of the Jodrell Bank radiotelescope, British astronomy (notably the Isaac Newton telescope) and the Daresbury electron synchrotron NINA. But their review of CERN is more important as it may affect real policy in two ways: first, a British scientific committee (see box) is studying whether Britain should remain a member of CERN, given the present £27-million price tag; and there are indications that Martin and Irvine's much-improved approach to scientific assessment in the CERN report is finding favour among British policy-makers, and could be applied to other fields. (A study has already been made of protein crystallography, for example, and delivered to Sir David Phillips, chairman of the Advisory Board for the Research Councils, which allocates basic research funds among the different research councils in Britain.)

But does the Irvine and Martin method work? The two researchers' earlier reviews drew criticism that in assessing scientific productivity (or the lack of it) in terms of indicators of publication volume, citation frequency and peer ranking of laboratories, they ended up with too bleak a picture of what began to seem "targets" rather than mere subjects for research. There were mitigating circumstances, and forms of productivity unmeasured by the indicators.

But Martin and Irvine's present effort — two papers published in *Research Policy* — can hardly be criticized on those grounds. The CERN survey is sophisticated, and the two investigators clearly recognize a development in their technique. For example, they previously described their method as one of "converging" indicators, thus begging the question of whether or not the indicators do indeed converge on a single black-or-white assessment of a laboratory. In the case of CERN, the indicators clearly do not converge, and, write the researchers, "the results based on each indicator need to be carefully interpreted". Interpret they do, basing their analysis on 182 interviews with particle physicists from the Soviet Union to the United States.

High-energy physicists, even those at CERN, who have seen draft versions of the papers describe the analysis as fair, balanced and (mostly) accurate. CERN is admitted to have just missed the boat on many occasions. The two quarks called charm and bottom, which highlighted a revolution in physics in the 1970s described by Martin and Irvine as "the American Revolution", could have been found at CERN, the study claims. Missing the discovery of these particles was described by one physicist interviewed by Martin and Irvine as "a tremendous failure".

CERN even produced serious errors. One particle, the *A2*, was claimed — falsely, as it turned out — to be split, or twofold, creating a puzzle which absorbed theoreticians for nearly five years during the late 1960s. But similar errors were made in the United States (the "high y" anomaly of Fermilab which caused equal consternation in the 1970s). Here the balance sheet is even.

In terms of positive results, CERN's first big machine, the 28 GeV proton synchrotron(PS), though on line before the equivalent alternating gradient synchrotron (AGS) at Brookhaven, was nowhere near as productive at the beginning. Physicists attribute this, above all, to lack of experience in the design of high-energy physics experiments in Europe. Then the intersecting storage rings (ISR), designed to collide PS protons with each other head on, was a brilliant technical success — but failed to produce revolutionary physics. (It helped in understanding the quark structure of nuclear particles, but could have seen and not missed charm and bottom.) Martin and Irvine suggest the machine was underexploited by CERN management. The 400 GeV super proton synchrotron (SPS) that followed came too late — more than four years after Fermilab in the United States had creamed the energy region — and had to rely on tidying-up (which, nevertheless, it did well, with resources that were considerably greater than Fermilab's).

Dr Herwig Schopper

The lesson, if there is a single one, is that Europe's laboratory was run too much by committee, whereas American laboratories owed more to individual flair. (Maurice Goldhaber at Brookhaven, R. R. Wilson at Fermilab and Pief Panofsky at Stanford Linear Accelerator Center are singled out.) But Martin and Irvine say that things are changing at CERN under director-general Herwig Schopper and that a more individualistic style has emerged.

Nevertheless, CERN has committed itself now to just one future machine and its 15-mile tunnel, the electron accelerator LEP now under construction. A future and more controversial report by Irvine and Martin will discuss the prospects for LEP and at lease one official involved in the British decision to support it has said he wishes he had had the report before him when he said yes.

Robert Walgate

POOR MARKS FOR ENTERPRISE?

V.F. Weisskopf

The report of the social scientists Ben Martin and John Irvine on the achievements of CERN, The European High Energy Laboratory in Geneva, was commented upon in "Nature" on 6 September 1984 under the title "Poor Marks for Enterprise". An observer well acquainted with the American and European activities in Particle Physics finds this report unfairly biased against the achievements of CERN.

European High Energy accelerator research on a large scale started only in 1960 with a completely novel social experiment, when the international laboratory in Geneva was ready for exploitation. Almost no tradition in "big science" existed at that time in Europe and certainly no experience of international scientific management existed anywhere. Nevertheless, within less than 24 years CERN had developed into the leading laboratory of the world in its field.

Why then do Martin and Irvine consider CERN's performance disappointing? They concentratre their attention on sensational peak discoveries. In this respect CERN contributed only during the last decades. But science does not consist of discoveries that make headlines in newspapers. The bulk of science consists of a broad front of painstaking investigations, leading to ever deeper insights. They are the soil from which some of those striking discoveries can develop. In this respect CERN has made many fundamental contributions to which Martin and Irvine give only passing attention.

It is true that some of the "peak" discoveries, in particular the so-called c- and b-quark, could have been found at the intersecting storage rings (ISR) at CERN. Other "sensational" discoveries, the violation of parity, the Ω-minus, and the two types of neutrinos, which CERN was blamed for missing, were made in the US before or at a time when research had barely started at CERN. It is certainly misleading to use only the number of those well-advertised successes as a measure for the quality of an institution. But even if one insists on counting the lucky events of that type, CERN did not fare so badly during the second half of its research period. It discovered a new form of radioactivity, the so-called neutral currents, the well-advertised quanta of the weak interaction, and probably even a sixth kind of quark.

Martin and Irvine ascribe the latest achievements to a recent change of spirit due to the present director. True enough, Herwig Schopper vigorously supported the last phases of the conversion of the SPS to a proton-antiproton collider and the subsequent successful research. He deserves every credit for it. However, Martin and Irvine do not seem aware of the fact that the conversion of an accelerator and the preparation of such giant experiments take much longer than the three and a half years of Schopper's directorship. The daring spirit and the technical inventiveness that have lead to these achievements date from much earlier times. They are a proof of the opportunities given to individual scientists, which Martin and Irvine considered as too restricted at CERN. They also show a steady growth of this institution towards its present world supremacy, that started long ago.

The comparison with the Fermilab in the USA is most unfairly presented. True enough, a number of great discoveries were made at that laboratory, such as the existence of a fifth kind of quark. But it was the CERN accelerator that was able to produce sharp and intense muon- and neutrino-beams with which the so-called structure functions of protons and neutrons were determined with an astounding accuracy. Moreover, the important "EMC-effect" was discovered, which has lead to a new view of the role of quarks in the atomic nucleus. One of the main purposes of those accelerators was the production of usable secondary muon and neutrino beams. In this CERN was much more successful than the Fermilab inspite of the latter's four-year lead. Why did Martin and Irvine not appreciate these successes and why was this considered by them to be dull physics?

CERN's contributions to the physics of nuclear structure are almost completely neglected in Martin and Irvine's report. CERN has the most advanced isotope separation device (Isolde) attached to its synchro-cyclotron where many new unstable nuclei were discovered and studied. Recently a low-energy source for antiprotons was constructed (LEAR) which provides unique opportunities for the study of antiproton interactions.

The success of CERN is based to a large extent upon a fact that was surprising for people who have observed the development of physics in the last half century. America was assumed to be superior in engineering and instrumentation as demonstrated by the development in the 1930's of cyclotrons, synchrotrons, linear accelerators, bubble chambers, etc. But the quality of the engineer-physicists at CERN turned out to be at least as high. They succeeded in constructiong accelerators of unusual reliability and flexibiliy under the leadership of the late Sir John Adams. Some people ascribed this to an over-conservative design, but it helped greatly in the exploitation of the machines. For example, it allowed the luminosity of the intersecting storage rings to reach more than ten times the design value and it facilitated the construction of the previously mentioned intense muon- and neutrino beams. Finally, it made possible the fast conversion of the 400 GeV accelerator into a proton-antiproton collider. The success of this conversion depended upon one of the most imaginative devices invented and carried out at CERN by Simon Van der Meer. It is the so-called stochastic cooling of the antiproton beam, a method that is going to be used in future colliders everywhere in the world.

Furthermore, the CERN engineer-physicists led by Kjell Johnson constructed the intersecting storage rings at a time when the American laboratories did not dare to take on this difficult task. Indeed, the first two colliders for protons were designed and constructed at CERN: The ISR and the proton-antiproton collider in the SPS tunnel. With these accomplishments CERN has pioneered a new style of experimentation which now is taken over by other centres of particle research. All over the world the newly planned High Energy devices are going to be such colliders. CERN continues the tradition by constructing the large electron-positron collider "LEP".

The Martin-Irvine report does not make much of these achievements; it all but ignores also the numerous new instrumentation ideas that were spawned at CERN, such as the multiwire proportional chamber and the vacuum ultraviolet imaging device. Both innovations are not only significant advances in detecting particles, they have important applications in medicine and in other sciences.

High Energy Physics is an international endeavour, more so than most other scientific activities. There is a constant exchange of physicists and engineers between CERN, the United States and other countries. It does not make much sense to consider the progress in that field as a competition between two continents. It is a collective achievement of all participating nations. The CERN effort has contributed essential ingredients. Particles physics would not be what it is without CERN's impetus, without its technical innovations and without its studies and discoveries, even if some of them did not reach the newspaper headlines.

Last and by no means least, CERN represents the United States of Europe in fundamental physics. It became an active symbol of the spirit of European unity. That is no mean achievement. It serves as the University of Europe for High Energy Physics and brings together the many individual Universities in Europe in scientific collaborations of unprecedented magnitude and complexity, frequently involving also Universities and Institutes of the U.S., of Eastern Europe and Asia. In spite of the obvious difficulties which such international collaborations entail, CERN has developed into the foremost High Energy Laboratory of the world. This is strikingly confirmed by a recent American Panel report on New Facilities for the U.S. High Energy Physics program, in which it was stated:

> "... The European facilities frequently provide a better level of support and/or a larger number of opportunities than the American equivalents."

How can such an institution be given "Poor Marks for Enterprise?

BEST COPY AVAILABLE

A Critique of Irvine and Martin's Methodology for Evaluating Big Science

John Krige and Dominique Pestre

> *Question:* Why do you spend money on high energy physics when you have so many other things to do?
>
> *Answer:* (1) According to Marx and Engels every discovery will have a practical application some time;
> (2) This kind of science presents the greatest possible challenge to engineers, physicists and industry, which is important to us;
> (3) Any civilization of any repute must be interested in the structure of matter, the origin of life, and the origin of the universe.
>
> (Response of a Chinese colleague to a question asked by V. Weisskopf.[1])

The work of John Irvine and Ben Martin (IM) deserves serious attention. The range of their studies is wide, they address themselves to issues of importance, particularly to science policy makers, their output is considerable, and they have made a marked impact not only in the scientific community, but also well beyond. Yet impact is not the same thing as quality, as IM would readily agree. And though there is much to recommend in IM's work; its value is, we believe, vitiated by a number of methodological inadequacies. It is on these that we wish to focus in what follows. The paper is divided into three main sections. In the first we summarize briefly the main stages followed by IM in presenting their results. We then go on, in the core of the paper, to criticize several aspects of their methodology. In conclusion we try to assess the value which their work has, despite the limitations we have identified.

Rather than trying to cover the whole corpus of IM's work, we base our discussion primarily on their recent studies on high-energy physics, and CERN in particular.[2] There are two reasons for this.

Social Studies of Science (SAGE, London, Beverly Hills and New Delhi), Vol. 15 (1985), 525–39

Firstly, our main interest is in their methodology, which is adequately represented in these papers. Secondly, as we are currently engaged in writing the history of CERN, we are able to draw on what we have learned to expose some of the limitations of IM's approach. While we recognize that, in making this choice, we may be suspected of wanting to defend CERN,[3] we wish to stress that that is not our concern. If we are 'defending' anything, it is intellectual rigour.

IM's Argument: The General Structure

The plan generally followed by IM when presenting their results can be summarized as going through four successive stages. To avoid misrepresenting their views we rely heavily on quotations from their work.

First Stage: A Critique of the Peer-Review System

IM stress the need 'for improved and more open methods of evaluating research performance in basic science' (*EBS*, 2). Typically, they give three main reasons for this.

(a) There has recently been a levelling off and even a decrease in the science budgets in industrialized countries. Thus 'existing financial commitments must generally be reduced in order to free the funds to support promising new research areas and young scientists [. . .] not a task for which peer-review has proved particularly effective'. (*EBS*, 3)

(b) With the concentration of resources in a few central facilities having annual budgets 'now running into tens or even hundreds of millions of dollars' there is a need 'not just for accountability to scientific peers, but for wider public accountability'. 'One solution to this problem lies in the greater use of output indicators in helping regulate the scientific system'. (*EBS*, 3)

(c) There has been an 'entrenchment of particular interests in decision-making bodies' reflecting 'previous patterns of resource distribution'. Instead of allocating resources on scientific merit, in the 'free market' of scientific ideas, a situation of 'oligarchy' and 'rigidities' has developed, which place an 'increasing strain on the peer-review system'. (*EBS*, 3–4)

BEST COPY AVAILABLE

Second Stage: How to Evaluate Basic Research?

IM's method relies on four main considerations. First, 'it is based on an input-output approach [. . .]. For basic science, some simplification is possible since one can concentrate on the primary output of contributions to scientific knowledge'. Second, the 'approach is institutionally focused'. Third, the 'approach is comparative, with the added condition that one can only legitimately compare "like" with "like"'. Finally, it 'involves the combined use of several indicators' (*EBS*, 5–6) like publication counts, total citations, number of highly-cited papers, peer rankings, peer comments . . . If the indicators converge (that is, when they 'all point in the same direction') IM 'regard the results of the evaluation as being relatively reliable, and certainly as being more reliable than those based on a single indicator like peer-review'. (*EBS*, 9)

Third and Fourth Stages:
From Past Performance to Future Prospects

The method is then applied to the 'past performances' of the facilities under study (stage 3). From this IM try to assess future prospects (stage 4). To this end they identify in the course of interviews with top scientists 'the factors that had structured success and failure in the past', and analyze 'further the various factors to ascertain which were likely to continue to exert an influence on research performance in the future [. . . and . . .] any new factors that are likely to emerge'. (*EBS*, 23) On this basis, they derive a set of comparative criteria to assess the prospects for the various facilities. And they draw their conclusions.

Can One Reasonably Expect to Evaluate Correctly Research Centres in 'Big Science' — to Improve Science Policy Decisions — Using Only Their 'Scientific Contribution to the Progress of Knowledge'?

(a) To avoid fruitless discussion, we want to stress that we do not doubt that IM *say* that outputs other than contributions to scientific knowledge are important.[4] And we see much to recommend their decision to concentrate on this one aspect, particularly in their recent work. What we object to is their insistence on calling these other outputs 'secondary reasons' which (in the case of high energy physics, for example) 'cannot explain why, over the last decade, nations have spent around $1000m a year on the subject'. (*EW*, 299) And to the fact that, whatever they may say, IM use only the one criterion, as measured by number of papers, citations . . ., to draw their conclusions.

(b) In our view it is quite wrong to assume a priori that considerations other than that of scientific merit are of secondary importance in deciding whether to fund Big Science. The quotation we have put at the head of this paper makes the point. Results obtained in our own studies confirm it.[5]

To illustrate, let us start with a simple historical question: Why was CERN established at the beginning of the 1950s? What were the reasons which convinced governments to give so much money to a scientific institution? We will not enter into details, but a crucial reason was the brain drain. Of course the physicists wanted to do scientific research and to compete in the production of knowledge with their American counterparts. But a huge facility working on the frontiers of research in physics was necessary in Europe to retain the brightest people in a situation of open (scientific and economic) competition with the USA. In other words CERN had — and perhaps still has — an important socio-economic function to fulfil.

Thus, without pretending to be exhaustive, some questions cannot be systematically relegated to secondary importance, questions about:

- manpower training (To what extent do big science facilities train people for industry? Do they train them in such a way that they have rather 'unique skills?[6]) What would be the consequences of closing such facilities down, or closing *it* down, when only one exists in a particular scientific field in a country? . . .;
- technological fall-out and other economical benefits in industry (reduction of production costs, development of new products, improvement of market image . . .)[7];
- long-term indirect effects on military technologies;[8]
- social role in testing new kinds of equipment, pushing them to their limits;
- national prestige.[9]

In sum, and even if it is difficult to assess this, 'Big Science' (or at least particle physics and space technologies) have a kind of role analogous to that of military research in the economical field.

(c) We are not claiming that 'non-scientific' considerations, rather than the production of scientific knowledge, are 'primary' or 'more important' in assessing Big Science. Our point is, rather, that it is wrong to weight them aprioristically and, correlatively, that the weight in fact given to these different outputs depends crucially on the *socio-historical context in which the decision is to be made*. At one extreme, if one has several institutes in the same country doing roughly the same work in a particular field, their contributions to scientific knowledge, as measured by IM, would conceivably carry the greatest weight in the minds of policy makers assessing their value. On the other hand, the 'bigger' the sciences and the fewer the number of facilities devoted to it, the less important this factor is likely to be. At the limit, if you have a facility like CERN, the only one of its kind in Western Europe, it is inconceivable that one could close it down just because its scientific performance is not the very best. Of course, it is reasonable today to demand of CERN that it does science which compares favourably with that done in analogous centres elsewhere in the world — though this would have been a ludicrous demand in 1960, for example. But this being achieved, the real problems begin: weighing a variety of outputs which cannot all be 'measured' in the same way; weighing them against the inputs; and finally weighing them against other outputs which may derive from similar investments in other fields of basic or applied research. In addition, these weightings must be done on a very long term basis because, while it is easy to stop a big centre, it is a complex process to restart it. As IM have shown, for CERN, it is only after 15 to 20 years of existence that the laboratory began to compete with its American counterparts. And this would be true for most facilities in 'Big Science'.

The Transition from 'Past Performance' to Assessing Future Prospects

(a) The first thing to notice is that the part of IM's work dealing with future prospects relies mainly, if not entirely, on peer-review, since the potential scientific outputs and the factors which structured, and are likely to structure, the functioning of research facilities in the future are derived from interviews. This implies:

— that IM rely essentially on the *average feelings* of the scientific community, leading us to doubt whether they can produce assessments substantially different from those obtained through a more classical panel system;

— that the results obtained using 'quantitative' indicators have little value for policy makers. In fact, what do these indicators measure? 'The Scientific Performance of the CERN Accelerators', as stated, for example, in the title of one of IM's articles? We do not think the answer is so simple. These quantitative indicators give at one and the same time an idea of the skills of the experimental teams, of the impact of the science policy of the institution, of the way people work (cultural styles or preferences), of the fact that, in the case of CERN, delays have occurred in building the machines because of the attitude of some Member States in the past, of good and bad luck, and so on. In other words, these indicators give a *global measure of the effects of these 'explanatory' factors on past scientific output*. To disentangle them, to weigh their importance one against the other, IM *need* to refer to peer interviews. Unfortunately, for policy decisions, it is this explanatory level which is crucial, and not primarily the global measure.

(b) Furthermore, when 'projecting' forward the factors which structured and are likely to structure the future functioning of the facilities they study, IM tend to rigidify them. Let us take an example. An interesting factor mentioned by IM — and one taken for granted by them and by the physicists themselves without careful assessment using historical data — is the opposition between 'a more bold, speculative ethos of physicists in the United States, where the research system was regarded as normally allowing more scope for individual initiative' and 'a more solid, safe and conservative approach to their experiments' by European high energy physicists. (*CII*, 272) This opposition, when identified through interviews on the past performance of CERN, appears as one of the reasons most frequently given to account for the small number of key discoveries made in the European laboratory. As key discoveries are *de facto* considered in IM's papers as essential to assess the contribution to knowledge, the American 'values' are connoted positively (and CERN is thought to have to 'change' in this respect). Hence the rigidity. In fact, what appears as an unnecessarily conservative and expensive way of doing physics may, in another scientific context

BEST COPY AVAILABLE

which is not easy to foresee, turn suddenly into a positive and perhaps decisive advantage, even in terms of producing key scientific results. This seems for example to be the case in the more recent successes of CERN (transformation of the SPS into a $p\bar{p}$ collider and subsequent discoveries of the Z_0 and Ws). As Clemens A. Heusch, of the University of California, Santa Cruz, put it:

> CERN's facilities have been designed and built by well-paid engineers [. . .] The ensuing high quality of machine building has paid off handsomely, only by the high standards of magnet and vacuum chamber construction can the success of converting the SPS accelerator into a colliding $p\bar{p}$ machine be explained.[111]

The point to stress here is how cautious one must be when using such explanatory 'factors'. First they must be studied critically in the light of *historical data*. If not, one is reduced to repeating general and banal statements. At this level of generality (see *CII*, 271, Table 9), nothing is really false, nor really true. But when trying to use them in the future, additional caution is required—fundamentally because a radical unknown exists.

(c) This leads us to the notion of risk linked with the use of advanced and highly sophisticated technology. When studying the past, one tends to underestimate the problem. If something has succeeded, with the wisdom of hindsight the risk is thought to have been 'reasonable'; if something has failed, the risk is often considered as having been 'too big'. When considering the future, it must first be recognized that an unknown is present and that, even without being deliberately biased, the experts often disagree on the importance of the risk and on the scientific consequences which will derive from it. In such situations there cannot but be a non-rational component in decision-making, a leap of faith. And the kind of 'objective data' which IM provide do little, if anything, to place such decisions on a more rational basis. What are we to make of statements as general as

> At Brookhaven [the problems . . . with Isabelle] were far more severe and were largely responsible for the delays to the facility's' construction programme [. . .]. The technical difficulties faced by the CERN accelerator builders [. . . $p\bar{p}$ collider] have also been extreme [though the situation] rapidly improved [. . .] over the following two years [. . .]. In the case of the electron-positron colliders, the technical problems associated with the SLC are felt by most physicists to be particularly severe [. . .]. (*CIII*, 32–33)

Perhaps this information is interesting for the layman. But for scientists and for policy makers it is far too imprecise. Depending on the importance of the decision to be taken, they need better documented and contradictory opinions. The 'ecumenical synthesis' derived from the views most commonly held among peers, and presented under the guise of objectivity, will be of minor interest — unless one is simply searching for reasons to 'justify' a decision already taken.

What about IM's Remarks about the Rigidity of Peer-Review?

To begin with, some precision is called for. When speaking about peer review are we speaking about peers asked to propose priorities between facilities in the same field (that is, comparable in terms of 'like' with 'like') or between facilities which 'produce' knowledge in different fields (say, high-energy physics and biochemistry)? In what follows we shall distinguish between these two kinds of choice.

(a) Concerning a choice between 'like' and 'like', the method developed by IM applies. What is more, as they say, the results obtained by asking peers to rank facilities have been, up to now, in rather good agreement with the results obtained using 'quantitative' indicators. In other words, they demonstrate that peers, when asked to compare 'like' with 'like' in a systematic manner, are rather reliable.

Is this last fact surprising? Contrary to IM, we would not lay such stress on the 'subjectivity'[11] of peer-ranking compared with the more objective results obtained through quantitative indicators. In fact, the data (particularly citations) used by IM are themselves the result of scientists' opinions about the works of their colleagues. They embody competent scientists' knowledge — though a 'knowledge' not obtained in a systematic manner — of which institution produced key discoveries and which gave systematic and reliable results. These scientists are likely to have a good general idea of what the results of IM's bibliometric measures are likely to be. In other words, the rather good agreement found by IM is not to be wondered at, and does not require special explanation. It is only if a disagreement were to appear that a special explanation would be called for.

(b) Turning now to the second case — the choice between different fields — real problems arise. As Weinberg has put it, 'to the expert in oceanography or in high-energy physics, nothing seems quite as important as oceanography or high-energy physics.'[12] But here, the method proposed by IM is of no direct use. If the willingness to solve this kind of problem through a 'rational' process of comparison (rather than through simple hard bargaining) has any sense, IM propose no solution. Or, at best, they hint that *the* solution is to cut the budgets of institutes performing badly on scientific grounds.

(c) It should be added that, when peers are asked to make judgements of scientific merit across sections of the same discipline (physics as a whole for example), they are not always that 'unreliable'. A recent study of the rankings made by a jury of physicists active in different fields of physics who were asked to assess proposals for basic research concluded (on the basis of bibliometric data) that 'the juries are competent: they know who is good and who is better' and their 'judgement has predictive value'.[13]

(d) Whatever its limitations, it does not follow that what IM do is of no interest. As they say, their approach 'provides means to help keep the peer-review process "honest"'.[14] However, it would be wiser not to let people think that their method can do more than this, it would be wiser to avoid the easy rhetoric we find in the First Stage of their presentation (see above). Doubtless rigidities exist in the peer review system, and some powerful oligopolies have developed. However, IM themselves 'show' that these are not biasing peers called on to assess the scientific merit of facilities in their own field. If such rigidities are important, it is precisely because they impede allocation of resources to new fields. While this provides IM's rationale for producing a more 'objective' and 'open' method of assessment, in fact the limitations of the methodology they develop preclude them from making a constructive contribution to the issue.

Further Less Important Criticisms

Let us agree that IM's main purpose is to provide 'a means to keep the peer-review system "honest"' (at least as far as these peers are asked *only* to give advice on the *scientific* performance of *existing* institutions working in the *same* field). That granted, our general feeling is that, *these boundaries having been established*, their method is worthwhile. However, some minor comments remain to be made.

(a) The need for rigour in comparing 'like' with 'like'

A cardinal tenet of IM's methods, and one with which we would agree, is that it can be used only to compare 'like' with 'like'. However, IM often play fast and loose with this notion. For example, in their comparative study on accelerators in East and West, they begin by distinguishing five kinds of centres equipped with five kinds of machines. When comparing publication output, they remain basically within these groups (*EW*, 304–06); their peer-review data, however, cut right across the distinction, ranking different kinds of facility on the same 'linear' scale, and so treating them as alike. Our point here is not to decide *when* they are right. We would just like more consistency and clarity.

Anyway, the decision about what is comparable is not just a matter of mere convenience. Mulvey wrote recently that if SPEAR (the electron-positron collider at SLAC) 'was responsible for the major part of the difference in score on which [IM] hang their case', it is because this kind of machine was 'the best type of machine for the physics of the 1970s'.[15] We do not know if he is right, but if he is, his statement reveals that *the definition of what is comparable must be clearly discussed in the light of each historico-scientific context*. What is more, this context will not be easily foreseen beforehand. As Mulvey put it, 'in research of this nature, until the experiments are done, no one knows which is the best choice'.

On other occasions also, IM's criterion for 'like' with 'like' is used too loosely. For example, in the same paper (the East-West comparison), IM explain clearly that the mechanisms of publishing in the Eastern and Western blocs are very different, and that the data they produce are not readily comparable. That granted, their quantitative results derive what conviction they have only because the differences between East and West are very big. They certainly add no rigour to what we already know or to the impressions gained by scientific peers.[16] More fundamentally, would it not have been more realistic simply to say that, given the great uncertainties surrounding quantitative measurements of publication outputs in the Eastern bloc, such comparisons have little to contribute to our understanding of the East-West difference?

(b) The notion of Partial Converging Indicators

Of course IM's response to the above would be that publication output is only one of several admittedly limited indicators. Though each is imperfect or partial in itself, the 'convergence' of the separate indicators gives evaluative results which are relatively reliable. Our main concern here is with the significance of the claim that the quantitative indicators 'converge' among themselves, and with peer-review. In our view it is misleading to describe the relationships which are involved in this way. Let us explain.

To begin with, what do the 'quantitative' indicators — publication counts, number of citations per paper, highly-cited papers, and so on — measure? To a first approximation, each of these indicators measures a different *facet* of the scientific performance of an organization (as compared with others), each contributes to the building of a *global* and *synthetical picture* of the organization in question. Thus CERN, for example, emerges as being at once productive (publication count), very reliable in its results, but weak in producing key discoveries (citation count). This picture is obtained by *adding together* the results of more or less parallel 'indicators'. That granted, the question of whether or not these indicators 'converge' is ill-conceived, and beside the point.

What of the relationship between the results of quantitative indicators and of peer review? Here the notion of convergence can gain some purchase. But it is primarily a 'convergence' between two overall pictures, one built by 'assembling' the quantitative data, the other obtained by interviews with peers. Here one can speak of these *images* as converging (or diverging). Yet even then the terminology is misleading, for it suggests that the two pictures of the organization are derived from *independent sources*. As we pointed out above, they are not, and it is not surprising that, until now, the pictures obtained in both ways are roughly identical.

Some Remarks on IM's Rhetoric

Our remarks in this section are inspired by the recognition that IM's papers are generally — and were for us — very convincing when read for the first time. While all of us use certain 'ploys' when writing to make our case, we believe that IM use a number of 'rhetorical rules' in a rather systematic way. We identify three in what follows.

(a) *First 'Rule'*: envisage all possible objections first, and then simplify. In this way, the reader's critical faculties are disarmed.

The examples of this 'rule' are numerous. It operates, for example, when IM present the various kinds of outputs one can consider; or when they stress that citation counts measure the impact of a paper, but not its quality or its importance; or when they show the limitations imposed by their data (East-West comparison); and so on. In general these discussions are well done and are accurate and stimulating. However, the problem remains that, having discussed all the limits, IM do not always take them into account during the demonstration and in their conclusions and summaries.

(b) *Second 'Rule'*: select one from a set of factors relevant to assessment, and then declare it the most important one.

To illustrate what we mean by this formula, we will take an example. In *La Recherche*,[17] IM consider the notions of quality of a paper ('well-done work, absence of mistakes' . . .), importance ('*potential* influence' if 'communication was perfect' — and, we would add, if the scientists knew clearly which article would be considered in the future) and impact ('effective influence'). But instead of saying that quality and importance *cannot* be measured — granted that they have a meaning — IM try to show that impact is the only important thing to know. They write:

> [T]he impact of a publication describes its *effective* influence within its field at a given moment, and so on the advancement of scientific knowledge. It follows that the impact of an article, rather than its quality or its importance, determines the extent of its contribution to scientific progress — from which arises its interest for a policy choice.

In other words, instea[illegible]mply saying that citation counts will give us a rough idea of th[illegible]y/importance/impact of an article, and that we will use th[illegible]use citation counts are easy to handle anyway, IM try to s[illegible] (fortunately) what we can calculate is precisely the most inte[illegible] factor.

This example is drawn from the microlevel of IM's work. The same tendency to elevate the specific feature they wish to discuss from *one*, or *a*, factor in assessment to *the*, or the most important factor, also occurs on the macrolevel, as we showed above.

(c) *Third 'Rule'*: do not hesitate to appeal to common sense and general knowledge to strengthen your demonstration.

This rule operates, for example, when one reads what we have called the First Stage of IM's demonstrations (see above). Everyone who practises history or sociology suspects that people and groups tend to defend their own interest, that there are dangers when the same people are both judge and judged in a decision process, that lobbying is an important activity of scientists, and so on. All this is generally true. It does not follow — as IM show — that peers cannot be 'honest' when asked to rank facilities in their own field. However, common sense baulks at this conclusion, and the refinements in our understanding of the limits of peer review demanded by IM's results are not made. Received wisdom remains intact, uncritical and uncriticized.

What have IM contributed?

While what we have said above is consistently critical of IM's work, we wish to stress that there is much of value in what they have done. We would like to identify the following important and helpful contributions they have made.

(a) They have collected, organized and published a colossal amount of information about the scientific performance of big and expensive scientific institutes. In so doing they have provided an invaluable data base, containing both quantitative measures and scientific opinions, which no policy maker or serious student of such institutes can afford to overlook.

(b) They have shown that when peers are assessing their own fields they can be reliable judges of scientific performance. And they have also shown the value of relatively cheap and semi-mechanical alternative techniques which can be used by social scientists and others 'external' to Big Science to measure various aspects of an institute's scientific performance.

(c) Where choices have to be made in a field where several similar research units within a country (or a group of related countries) are competing for resources, they provide policy makers with sound information for assisting a rational decision.

(d) Finally, there are the 'spin-offs'. Speaking for ourselves, what they have done in the case of CERN is a source of inspiration for the historical work we are engaged in at the moment. Quite apart from anything else it will help us ask pertinent questions of our archival documentation.

• *NOTES*

1. This remark was made by V. Weisskopf at a seminar organized in conjunction with the Study Team for CERN History in September 1984. The original is on a tape in the CERN Archives, Geneva.

2. The core papers we are using, with the abbreviations we shall use to refer to them in the text are: (CI), Ben R. Martin and John Irvine, 'CERN: Past Performance and Future Prospects — I. CERN's Position in World High-Energy Physics', *Research Policy*, Vol. 13 (1984), 183–210; (CII), John Irvine and Ben R. Martin, 'CERN: Past Performance and Future Prospects — II. The Scientific Performance of the CERN Accelerators', ibid., 247–84; (CIII), Ben R. Martin and John Irvine, 'CERN: Past Performance and Future Prospects — III. CERN and the Future of World High-Energy Physics', ibid. (forthcoming); (EW), John Irvine and Ben R. Martin, 'Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators', *Social Studies of Science*, Vol. 15 (1985) 293–394; (EBS), John Irvine and Ben R. Martin, 'Evaluating Big Science: CERN's Past Performance and Future Prospects', *Scientometrics* (forthcoming).

3. Our concerns are thus different from those of V. F. Weisskopf and J. H. Mulvey, whose responses to a report on the work of Irvine and Martin appeared under the heading 'In Defence of CERN' in *Nature*, Vol. 311 (18 October 1984), 599–600.

4. They have explicitly considered them, in fact, in their John Irvine and Ben R. Martin, 'The Economic Effects of Big Science: The Case of Radio Astronomy', Proceedings of the International Colloquium on Economic Effects of Space and other Advanced Technologies, Strasbourg, 28–30 April 1980 (Ref. ESA SP–151, September 1980), 103–16.

5. Our work is available in a number of internal reports produced at CERN in the so-called 'CHS' series. See, in particular, D. Pestre, 'Préhistoire du CERN: le Temps d'un Optimisme Raisonnable, décembre 1950 – août 1951' (Geneva: CERN, 1984), 13–14.

6. Irvine and Martin, op. cit. note 4.

7. The thesis of H. Schmied, *Essai d'évaluation des Effets Economiques des Conventions Conclues Entre le CERN et l'Industrie* (Strasbourg: Université Louis Pasteur, 1977), is a good example of a study of such questions. While his method has its limitations, he is rather convincing when showing that in some cases, the ratio of the economic utilities produced in industries by the contracts with CERN to the amount of the contracts themselves is far from being negligible (31.6 in the case of precision engineering; 17.3 in the case of computing) or when he argues that roughly 80% of the total utility reported by firms derives from sales to clients foreign to high energy physics and nuclear physics. See also J. Grinevald, A. Gsponer, L. Hanouz and P. Lehmann, *La Quadrature du CERN* (Lausanne: Editions d'en bas, 1984), 123–39.

8. Consider, for example, President Reagan's recent 'Star Wars' speeches. See also Grinevald et al., op. cit. note 7, Chapter 1, for some suggestive remarks about military linkages.

9. IM themselves point out that, in the Eastern bloc, 'matters of national prestige' have played a major role in the funding of high-energy physics (*EW*, 301).

10. Clemens A. Heusch, 'US Participation at CERN: A Model for International Cooperation on Science and Technology', CERN EP Note, 20 January 1984, 11–12.

11. See particularly John Irvine and Ben R. Martin, 'L'évaluation de la Recherche

Fondamentale est-elle Possible?' *La Recherche*, Vol. 12, No. 128 (December 1981), 1406–16.

12. A. Weinberg, 'Criteria for Scientific Choice', *Minerva*, Vol. 1 (1962), 159–71.

13. W. Westera and C. le Pair, 'Evaluation of a Science Policy Instrument', in W. Callebaut, S. E. Cozzens, B-P. Lecuyer, A. Rip and J. van Bandegem (eds), *George Sarton Centennial* (Ghent, Belgium: Communication and Cognition, 1984), 488–90.

14. Ben R. Martin and John Irvine, 'Research Evaluation: Why? How?', in Callebaut et al., op. cit. note 13, 241–43.

15. Mulvey, op. cit. note 3.

16. See also the section of the same paper entitled 'Accelerator Outputs — Scientific Publications' (*EW*, 304–08). Here IM identify four possible sources of bias in their data collection techniques used to compare publication counts. On the second of these, the difference in publication practices, IM are particularly convincing (*EW*, 307). Unfortunately the authors do not estimate the possible bias in this case, and in their conclusion they quietly shelve it, giving an overall estimate of bias which does not take it into account (*EW*, 308).

17. Irvine and Martin, op. cit. note 11, 1409–10.

John Krige and *Dominique Pestre* are both members of an international team currently writing the history of CERN. The latter is the author of *Physique et Physiciens en France, 1918–1940* (Paris: Editions des Archives Contemporaines, 1984). *Authors' address:* Study Team for CERN History, c/o CERN, CH-1211 Geneva 23, Switzerland.

Responses and Replies (continued)

Critical Remarks on Irvine and Martin's Methodology for Evaluating Scientific Performance

H. F. Moed and A. F. J. van Raan

Our article is divided into three main sections. In the first section we focus on Irvine and Martin's (IM's) method of 'converging partial indicators' for evaluating scientific performances. We first

---

Social Studies of Science (SAGE, London, Beverly Hills and New Delhi), Vol. 15 (1985), 539–47

summarize the essential elements of this method. Next, we state briefly what we believe to be the outcomes of IM's efforts to test their method. Next, we criticize IM's claim as to the validity of their method, and we discuss briefly our opinion on the application of the various indicators. In the second section we discuss IM's choice of the object of their method: 'Big Science' research facilities. In a further section we comment on a few technical problems related to citation analysis. Finally, we mention a number of valuable elements in IM's work, and we summarize our main criticisms.

With this article we respond in the first place to IM's recent publications in *Social Studies of Science*,[1] but in some sections we comment on their work as a whole.

Critique of the Method of Converging Partial Indicators

Short Outline of IM's Method

IM's method of converging partial indicators for comparing scientific contributions is described extensively in their paper on radio-astronomy institutes[2] and more briefly in their article on the Isaac Newton Telescope. Essentially, several 'measures of science' are applied, based on publication counts, citation counts, and peer judgments; these measures are assumed to be indicators of scientific progress. They are at best 'partial indicators', influenced by a network of interrelated factors, in which the size of contribution to scientific progress is only one. By combining the various indicators, and applying these to research groups, comparing 'like' with 'like' as far as possible, the influence of other, disturbing factors can, according to IM, be minimized. Then, only in those cases where convergent results are obtained, it can be assumed that the influence of these other factors has been kept small. Thus the convergent indicators provide a reasonable estimate of the relative contribution to scientific progress.

Tests of 'Appropriateness'

IM claim that they tested the 'appropriateness' of their method in a number of empirical studies, and that these tests have been successful. (See, for example, *INT*, 54.) What do IM mean by this term 'appropriateness': in other words, by what criteria were these empirical tests performed? They state that the appropriateness criterion of their method is the extent to which 'convergent results' can be obtained.

Performing such tests, one should demand that the method is described clearly and applied rigorously, and particularly that the phenomenon to be observed (that is, 'convergence') is defined adequately. However, this concept of convergence is poorly developed. It is not operationalized at all. In fact, in their recent 'East and West' paper, IM do not any more use the term 'convergence' *as such*, but they add predicates like 'some' or 'a certain' to it (*EW*, 299, 322). Even worse, in their paper on the Isaac Newton Telescope, the various indicators do *not* converge; but this does not stop IM concluding that 'the method appears to be capable of comparing the relative contributions of optical telescopes, and of producing significant results' (*INT*, 74). In our view, in order to perform these empirical tests adequately, one should use clear and illustrative techniques to indicate convergence. In the social sciences, especially in psychology, these 'metric' techniques are well known (multidimensional scaling, clustering techniques, and so on).

A second example illustrating the lack of rigour in description and application of the method relates to the concept of 'matched groups'. One of the criteria for matched groups appears to be that the groups concerned publish in the same journals. East and West high-energy physics groups do publish in different journals and should therefore not be considered as matched groups. Consequently, according to IM's methodological rules, the Eastern and Western groups cannot be compared. Nevertheless, IM do compare these groups. Another related point is the incomplete coverage of journals. In the East-West comparison, coverage of journals, especially for the East bloc, is incomplete. In their Table I (*EW*, 300), IM state that incompleteness of journal-coverage is not a serious problem in Big Science. But high-energy physics *is* Big Science! Why then do IM make such efforts to estimate the possible bias due to incomplete coverage?

However, we would like to make a point with respect to their work as a whole that seems even more important. One should distinguish between the validity and the applicability of their method. 'Validity' refers to the extent to which the concept of contribution to scientific progress, operationalized by a set of converging indicators, coincides with the concept of scientific progress which the method claims to measure. 'Applicability' of the method, on the other hand, refers to the conditions under which the method can be applied in order to provide valid results. One of the conditions for applicability of IM's method — in our view the strongest one — is the occurrence of convergence with respect to the various indicators used. The tests performed by IM partly do — and in principle can — examine the *applicability* of the method, but they do not and cannot provide any results relevant to the question of its *validity*. In particular, their tests can never provide results that invalidate their method, or even question its validity. If divergence should be observed in some cases, this result does not have any implications for the validity of the method. According to IM's methodology, one should then simply conclude that apparently their method does not work in such cases. So the question of the validity of their method has, in fact, disappeared in IM's studies. We will now discuss this problem in more detail.

The Problem of Validity

The question of whether IM's method — if applicable — actually measures what they claim to measure, is nevertheless extremely important. In our view, an inconsistency appears in their defence of the claim that convergent indicators do provide reasonable estimates of contribution to scientific progress. From the statement that all indicators are influenced by a network of interrelated factors of which the size of contribution to scientific progress is only one, we can conclude that if disturbing factors are minimized, convergent results should be obtained in combining the various 'partial' indicators. However, IM reach precisely the opposite conclusion: they conclude that if convergent results are obtained, the disturbing factors are (probably) minimized. They do not provide any evidence in favour of their conclusion. On the contrary, they seem to provide evidence against it!

In their papers, a number of processes or mechanisms are described that might lead to convergence of the various indicators, without any guarantee that disturbing factors affecting these indicators are minimized. First, the applied indicators appear to influence each other. Second, it is a plausible assumption that existing disturbing factors will affect several indicators in the same direction. We illustrate this by giving two examples. In the Isaac

Newton Telescope paper, we see that this telescope has produced very few scientific papers compared to the US telescopes considered. Differences of the same order of magnitude are observed for citations (impact). IM write:

> What factors give rise to these apparently very large differences in impact between the British and American telescopes? Undoubtedly the main factor was the substantially lower publication rate of the former. (*INT*, 67)

In the Radio Astronomy paper, IM state that the citation rate of a scientific paper is determined partly by its impact, and partly by social or political measures that, amongst others, 'can be expected to vary in a systematic way between individuals, or groups of scientists occupying different cognitive and social locations' (*RA*, 71). From the discussion on peer evaluation it appears that 'social and political pressures within the scientific community', and 'diversity in the cognitive locations of peers' (*RA*, 73) affect the validity of peer-evaluation measures as indicators of scientific progress as well.

In our view, using their own conceptual framework, one should conclude that convergence of IM's partial indicators is at best itself a partial indicator of minimization of disturbing factors. As a consequence, convergent results provide at best a partial indication of scientific progress.

Our View

We think that quantitative measures — both bibliometric indicators of output and impact, and ratings of peer judgment as well — can provide valuable empirical data for generating and testing questions and hypotheses on various aspects of research performance, rather than an instrument claiming to measure 'the contribution to scientific progress'. We would like to illustrate this view. At least in their two recent *Social Studies of Science* papers, IM actually seem to use their quantitative data in tests of hypotheses — although they do not state this explicitly, and although they still stress the importance of their method of converging indicators. For instance, in their 'East and West' paper, they essentially test a number of propositions made by emigré scientists and (bodies of) Western scientists on Russian scientific performance and science policy, presenting consistent empirical evidence that seems to confirm most of these propositions.

In their Isaac Newton Telescope paper, somewhat contradictory claims of groups of English scientists on research performance and research management with respect to the telescope are examined in the light of a large amount of empirical data. IM's results here are indeed significant, particularly because one of the claims does not seem to be supported very well by the indicators (divergence!).

This last example, especially, illustrates the particular use of quantitative measures which, in our opinion, is most valuable in research policy: namely, the application of quantitative indicators as a test of how 'local' peers judge the relative research performance of 'neighbouring' research groups in their field — for instance, in their own institution or in their own country — and of the criteria they use in allocating funds. The outcome of such a test is a conclusion about the extent to which claims of 'local peers' are confirmed or rejected by empirical evidence — according to the view of those performing such a test. Observed discrepancies should lead to questions addressed to the local peers, and to more detailed, qualitative analysis.

Critique of the Choice of the Object of Evaluation

IM deal with a rather high (and, at the same time, rather curious) aggregation level: not a research centre or a (very) large institute, but a facility, an accelerator, within a research centre. This choice of aggregation level has important policy-relevant consequences. As soon as, at least in the West, a specific facility — like an accelerator — loses its position at 'the front', scientists will move to other facilities. It is not then any longer interesting to follow the 'performance' of the obsolescent accelerator. It would, however, be relevant to policy to follow specific (groups of) scientists: in fact — at least again in the West — the participating university groups. *These groups form the basis of scientific progress* and are, in our opinion, the most interesting (and therefore most policy-relevant) 'level' to evaluate.

We can add a second point to this 'aggregation level' or 'object choice' problem. Surely no important country (US, USSR) or international group of countries (Europe: CERN) will close down, or even put on ice, large prestigious objects like high-energy physics centres, radio-astronomy centres, and so on, on the basis of evaluation studies. These centres are simply too rare, too prestigious. At most, national authorities will introduce changes in such a centre — for example, a very obvious intervention like closing down a particular part of the centre (a specific accelerator). Only when a relatively large number of expensive, similar centres are present in one country — like the various high-energy physics centres in the US and the USSR — may national policy lead to a gradual closing down of one of these centres. But in all these cases participating groups of scientists would already have moved to other facilities. They will — again, at least in the West, where scientists really can 'change facilities' — 'predict' such a national policy measure.

Some Important 'Technical' Problems

Let us finish with a few technical, but important points, concentrating mainly on IM's recent 'East and West' paper. First, it is not true that mastering the technical problems in citation analysis is impossible with computer scanning, as stated by IM. In their opinion, manual scanning is the best way. We do not agree. It all depends on the search-algorithm used in computer-assisted analysis techniques. We refer to our work.[3] In manual scanning important difficulties are involved, not the least being a very obvious factor like the scanner getting tired (or even crazy). We contend that sophisticated search-software is much more recommendable.

Second, IM discuss and estimate three main sources of likely bias against Eastern bloc accelerators in their citation data. We focus here on the first and the third. The first source of bias is that papers scanned by IM 'lose' citations from journals not covered by ISI. The third is that papers published in the main Soviet journals tend to contain less references, on average, than those in equivalent Western journals. IM do not explain why these sources would cause a bias particularly against Eastern bloc accelerators. In principle *both* Eastern and Western papers scanned by IM suffer from the citations that are 'lost' (first source) and from the shorter lists of references in Russian journals (third source). Why then should Eastern papers suffer more than Western papers (up to 20–40 percent)? IM seem to assume that (Eastern) papers published in Eastern journals refer almost exclusively to other Eastern papers (and *not* to Western papers). However, they do not provide any empirical evidence in support of this assumption. In our opinion, an analysis of the sources of bias should deal with this assumption — and, more generally, with the question of which specific citation patterns exist between the two blocs.

Finally, IM state that:

> The citation rate provides an indicator of the average impact per paper from each accelerator and *thus* allows for differences in the *scale* of research activity at each facility. (*EW*, 338, fn 52, emphasis added)

This is probably *not* true. In a recent study on characteristics of citation behaviour, we found evidence that suggests that the 'citations per publication' is affected positively by the number of publications, and therefore does not allow properly for differences in the scale of research activities.[4]

Concluding Remarks

We would like to emphasize our appreciation of IM's enthusiastic activities. The amount and the range of their work is impressive. Without doubt, IM have played, and still play, an important role in the development and application of quantitative-empirical analyses of research performance. Their work contains many valuable elements. They have stated firmly that one should not rely on bibliometric data in analysis of research performance: other types of indicators should be applied as well (for instance, peer-review indicators). They have clearly introduced the distinction between 'quality' and 'impact' of research performance — and, more especially, 'impact' as a theoretical concept, reflected more or less adequately in citation counts. In addition, they have defined clearly the concept of (partial) indicators. And we would finally mention, as a valuable element in IM's work, their effort — at least in their latest papers — to analyze the causes of the observed differences in research output, impact and peer-recognition. We firmly agree that, from the point of view of research policy, one should obtain insight into the causes of the observed level of research performance in order to make proper, effective policy decisions.

Nevertheless, our critique is evident. IM 'promise' too much. We believe that IM's method of 'converging partial indicators' can provide at best a *partial* indication of contribution to scientific progress, and nothing more than that. In addition, performance analyses of large facilities such as accelerators have a limited relevance for research policy. In our view, quantitative measures of science are particularly useful in tests of how 'local' peers judge the performance of 'neighbouring', or even their own, research groups and of the criteria they apply in allocating research funds. Observed discrepancies should lead to questions addressed to these local peers, and to more detailed, qualitative analysis.

• *NOTES*

1. J. Irvine and B. R. Martin, 'Basic Research in the East and the West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators', *Social Studies of Science*, Vol. 15 (1985), 293–341 (referred to below as *EW*); Irvine and Martin, 'Assessing Basic Research: the Case of the Isaac Newton Telescope, ibid. Vol. 13 (1983), 49–86 (*INT*).

2. B. R. Martin and J. Irvine, 'Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, Vol. 12 (1983), 61–90 (*RA*).

3. H. F. Moed, W. J. M. Burger, J. G. Frankfort and A. F. J. van Raan, 'The Use of Bibliometric Data for the Measurement of Research Performance', *Research Policy*, Vol. 14 (1985), forthcoming; Moed, Burger, Frankfort and van Raan, *On the Measurement of Research Performance: the Use of Bibliometric Indicators* (Leiden: Research Policy Unit of the University of Leiden, 1983), 1–199.

4. H. F. Moed, W. J. M. Burger, J. G. Frankfort and A. F. J. van Raan, 'Characteristics of Self-Citation Behaviour: References to the Oeuvre of a Research Group', in W. Callebaut et al. (eds), *George Sarton Centennial* (Ghent: Communication and Cognition, 1984), 472–73.

Author's address: Research Policy and Science Studies Unit,
Stationsweg 46, University of Leiden, 2300 RA Leiden, The
Netherlands.

The Case of the Disappearing Caveat: A Critique of Irvine and Martin's Methodology

Robert Bud

The work of Irvine and Martin (IM) has provided a wealth of statistical data to be taken into account by future policy makers and historians. Those analysts will interpret the data in terms of the significance of what has been counted. This need for careful interpretation was sensitively recognized by IM in the methodological introduction to their paper on progress in radio astronomy (which I will refer to as *RA*).[1] Even then they were to underestimate a fundamental problem, and its difficulties have been magnified as the sensitivity shown in the methodological introduction has been abandoned. The authors have failed to avoid the question: What is research in science for? That is a black hole into which their analysis is irretrievably sucked. First they impute unrealistic goals to science, and then they use indicators which do not indicate levels of achievement of those hypothetical goals.

This Response focuses on their recent paper on 'Basic Research in the East and West' (*EW*).[2] The object of that work is to evaluate the scientific outputs of high-energy accelerators in the Eastern bloc compared to those in the United States and Western Europe. But if one is to 'evaluate' science, or indeed anything else, performance must be matched to goals. When these are embedded in three such different cultural systems the issue is particularly salient.

The objectives of science are explored by IM themselves in their recent book, *Foresight in Science* (*FS*).[3] Here it is explained that the conventional differentiation of 'basic research', 'applied research' and 'experimental development' may be useful for statistical purposes, but not for science policy (*FS*,2). Within 'basic research' one can distinguish in terms of the patron's objectives between 'pure or curiosity oriented research' and 'strategic research' (*FS*, 4). The

---

Social Studies of Science (SAGE, London, Beverly Hills and New Delhi), Vol. 15 (1985), 548–53

former is promoted for no 'long term economic or social benefits other than the advancement of knowledge', whereas the latter is expected to 'produce a broad base of knowledge likely to form the background to the solution of recognised current or future practical problems' (*FS*, 4). The authors explain in more detail the character of strategic research. This is defined, they emphasize, by the interests not of the scientists but of the sponsors, normally large science-based firms, or governments (*FS*, 5). The objectives are: a) to contribute background knowledge required in the development of new technologies, and b) to develop links with relevant academic research communities which would enable the firm or government to exploit knowledge created through the efforts of others (*FS*, 5). Both emphasize the importance of developing understanding within the sponsored laboratory. Curiosity-oriented research is less carefully examined. It is in their analysis an empirically residual class, basic science for which no long term application has been considered (*FS*, 4). No historical justification is presented for the identification with it of a glibly described 'traditional notion of academic research carried out with the aim of producing new knowledge primarily for its own sake' (*FS*, 3). The history of the patronage of academic science would in fact suggest that whatever its 'real' utility, it has always been expected to have a variety of economic, cultural, educational and even political outcomes.

These categories at least offer parameters within which research can be evaluated. They provide a means for separating between the objectives of the scientists and those of their patrons. They are regrettably ignored in the 'East and West' case study. Although, given IM's differentiation, research into high-energy physics appears to be classifiable as strategic research, their evaluation assumes the quite different orientation towards knowledge for its own sake. It is admitted that the research 'also leads to substantial educational benefits (in the form of highly trained scientific personnel), to various types of technological "spin-off", and even to broader political benefits such as increased national prestige and improved international cooperation,' (*EW*, 299) but these benefits are written off as 'secondary' and trivial in an undocumented claim: 'Yet these secondary reasons cannot explain why, over the last decade, nations have spent around $1000m a year on the subject' (*EW*, 299). Nevertheless we learn later (*EW*, 301) of the belief that such research would contribute to national nuclear energy research during the 1950s and early 1960s. For the containment in time no evidence beyond the rhetoric of scientists is produced. The very same paragraph emphasizes the national prestige considerations lying behind the East European effort in general, and the building of the Serphukov reactor in particular. The emphasis is well justified by the survey of East European physicists relegated to footnote 73 (*EW*, 340). Although we are left in the dark about Western decisions, from Greenberg and others, we know that these have also been politically complex. Gibbons' study of the 300 GeV accelerator showed the kind of considerations underlying policy advice to a Western government during the late 1960s. 'Extrinsic' goals are shown there to have been important.[4]

All this lurking empirical evidence that accelerators have many local functions is ignored. IM establish their own criteria for evaluation, quite separate from those of participating patrons. We are told that the laboratories are to be evaluated according to their contribution to 'knowledge' (*EW*, 299). This perspective leads to a recurrent problem: 'knowledge' for them, unlike the fruits of strategic research, is not situated in any particular community. On the other hand the indicators of knowledge production are community based. That is clear from the initial explanations of the chosen indicators. But the author's self-consciousness about the constraints of their measuring system fades through the paper.

Two modes of evaluation are designated: contributions to scientific 'production' and to 'progress'. Neither can be measured; instead their magnitude is to be judged through the use of partial indicators. These are, we are specially reminded, 'mental constructs not to be confused with reality' (*RA*, 66, fn 11). Three are chosen in the radio-astronomy paper, and are used again in the study of high-energy physics: numbers of papers, citations, and peer assessment. The following cautious explanations are articulated in the initial analysis. Since it will be argued that the caution is later ignored, it is worthwhile noting the author's own qualifications.

Peer assessment is used as an indicator of contribution to progress directly. It is recognized that a complicated series of intellectual and social processes is involved in peer review. What determines scientists' opinions of their colleagues is recognized to be most unclear. Similarly, the use of citations is explained in a sophisticated manner. The authors differentiate between the 'quality' or 'importance' of papers, which are properties of the papers themselves, but are inaccessible, and 'impact', which describes the 'actual influence on surrounding research activities at a given time'

(*RA*, 70). It is only with the last that the authors are concerned. They claim no necessary relationship with either quality or importance. An elegant implication of this approach is that it takes into account all those social factors beyond some notional 'importance' of the paper. These may include, we are informed, the location of the author, and the prestige, language and availability, of the publishing journal. It is not an absolute factor, varying according to the 'cognitive and social location of the assessor' (*RA*, 70–71). Finally, the authors claim that number of papers can be used to indicate contribution to 'scientific progress'. This is done with caution since we are told that 'we have to consider why scientists publish papers, and to realise that they do so not only to present valuable results, but also for social, political and career reasons.' (*RA*, 66)

These indicators, it is claimed, can be used together to compare the contributions to 'scientific progress' of laboratories. Given their culturally sophisticated concept of 'progress', it is important that the laboratories compared operate within the same culture. As IM say:

> What is ideally required is that there should be two or more groups, working in the same specialty over a similar time period, publishing in the same journals, supported with a roughly similar level of resources, and situated in a similar institutional context. (*RA*, 75)

The moral underlying this methodological account is that statistical results have to be interpreted before convergence or otherwise can be concluded. Not only do raw data need to be analyzed before being turned into conclusions about the magnitude of the indicators: those indicators themselves need to be interpreted. Each indicates a different phenomenon. Only when the interpretations coincide can one properly conclude that they have 'converged', and on what they have converged.

In IM's 'East and West' study of high-energy physics, the care taken with the collection of data is quite remarkable. Perhaps only those with experience of such work can appreciate the assiduity represented in a few short tables. The following conclusions are reached: a) fewer papers were published based on work in Soviet reactors; b) these papers were less cited in eleven major international journals taken together; c) anonymous physicists thought poorly of Russian work. These are, individually, interesting results. However before we can say that they converge, we must ascertain what they individually mean.

Regrettably, the interpretation of the quantitative data is perfunctory. Conclusions are drawn only about overall magnitudes. The understanding of what may be happening is derived from the interviews, with all their problems. Descriptions of East European weaknesses are combined with explanations of them. All depend on the complex, and uninterpreted, interests of the commentators. We can only note, without understanding the implications, that the number of Eastern European scientists interviewed was an order of magnitude smaller than that of the Westerners; and that we are not told of the method by which the interviewees were chosen. This is particularly problematic, given the small number of East Europeans.

The analyses of publication counts and citations could have been richly suggestive, given the author's sophisticated understanding of publication and citation as social processes. However, the only suggested interpretation of Soviet publication counts is that one might expect that with less pressure to publish trivia the average impact would be higher. Citation data, despite its complex significance, is subjected to similarly slight analysis. As IM had emphasized, it indicates 'impact', which is determined by the relationship between producer and reception communities, mediated by language and a whole host of social and cultural phenomena. Only a peep at the implications is provided — by the interesting study of citations to East European work published in Western journals, which indicates a dramatic difference in average impact between papers from the Soviet reactors published in the West, and those published in local journals. This is explained merely in terms of objective 'quality'. Meanwhile the segregation of Eastern and Western scientific communities is not examined and its implications lie unexplored. The authors are also silent on the intellectual and social objectives of publication and citation in East and West. Instead there is a silent shift from the category of 'impact' to that of quality. There is reference to the poor 'record' of Serpukhov. The result is that although the authors claim to have found the convergence of three separate indicators, what they have is one detailed indicator (the interviews) whose results give some meaning to the statistical analyses of citations and production. We are not provided with independent access to the meaning of those tables.

The interpretive problem with all the indicators is the same: What are their implications for our understanding of the relationship between the scientific communities? Unfortunately, since the

BEST COPY AVAILABLE

authors have defined their interest as 'knowledge', or 'progress' worldwide, the accumulation of local expertise and the mechanism of its interchange are defined as irrelevant. The imputed objectives of scientific activity rule out realistic interpretations of the chosen indicators. The two problems highlighted in this Response — the imputed objectives and the uninterpreted indicators — are therefore interrelated.

The historical account is consequently bedevilled by two empirical problems: it is not clear what phenomena are indicated here; and we do not know what phenomena we, as evaluators, should be interested in. In view of IM's meaning of 'outputs', their conclusion that 'the scientific outputs from each Eastern-bloc accelerator have been small in comparison with the nearest equivalent Western facilities' (*EW*, 334), may be trivial.

• *NOTES*

1. Ben R. Martin and John Irvine, 'Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, Vol. 12 (1983), 61–90.

2. John Irvine and Ben R. Martin, 'Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators', *Social Studies of Science*, Vol. 15 (1985), 293–341.

3. John Irvine and Ben R. Martin, *Foresight in Science: Picking the Winners* (London: Frances Pinter, 1984), esp. 2–7.

4. Michael Gibbons, 'The CERN 300 Gev Accelerator: A Case Study in the Application of the Wienberg Criteria', *Minerva*, Vol. 8 (1970), 180–91; on American considerations, see Daniel S. Greenberg, *The Politics of American Science* (Harmondsworth, Middx: Penguin, 1969), 261–302.

Author's address: The Science Museum, South Kensington,
London SW7 2DD, UK.

The Possibilities of Science Policy

H. M. Collins

Irvine and Martin (IM) have been responsible for some serious questioning of British science policy. To date, the outcome of their work seems likely to be beneficial for those who seek some basic nourishment from the research budget, rather than the five-course dinners of Big Science. Nevertheless, their research is not without flaws, nor is it obviously the best way for science policy to make use of recent work in science studies. I will criticize it on two grounds. First, it misses the cognitive 'wood' for the institutional 'trees', because of the way in which it disaggregates science. Science policy ought not to be about maintaining efficient institutions, but about maintaining the sort of *cognitive* community that will produce the desired scientific products.[1] The second, interrelated, criticism is that their measures of scientific output are inappropriate, since they reflect only the internal reward system of science; optimum performance according to such measures is not necessarily optimum for science policy. My criticisms of their work on the particular topic of high-energy physics institutions[2] are implicit in these more general remarks.

IM have studied science by breaking it up into units of comparison defined by what I will call 'non-cognitive boundaries'.[3] Usually the boundaries chosen have been those of institutions — the laboratory, the university department, the discipline, the nation. The analysis has then turned on the comparison of output/input ratios of these units. The greater the output/input ratio the better — that is, more cost-effective — the unit. This is certainly an improvement on measures of output alone, and much better than the widespread practice of assessing ability by input alone (how much research money is attracted). However, to develop a policy with cognitive goals in view, it is essential to start by disaggregating science

Social Studies of Science (SAGE, London, Beverly Hills and New Delhi), Vol. 15 (1985), 554–58

BEST COPY AVAILABLE

according to cognitive rather than institutional boundaries — that is, to think of science as being made up of sets of research areas which involve scientists who interact, or mutually refer, *across* institutional boundaries, because of their common cognitive interests. The boundaries of such areas do not necessarily map on to the boundaries of institutions.

Turning to IM's measures of output — publication counts, citation statistics and peer-evaluation — these all reflect the reward system of science. This is probably why these 'converging partial indicators', converge so readily.[4] But the reward system of science has an 'atomistic' ethos. Prizes, citations, peer appreciation, and the accolade of having made a major contribution, do not go to all those who take part in a scientific endeavour, but to those who complete the work first. As Medawar says, 'Much of a scientist's pride and sense of accomplishment turns . . . upon being the *first* to do something',[5] and the reward system reflects this. The atomistic reward system *does* map on to institutions. That is why IM's measures of output appear to make sense. Institutional units can be readily compared in terms of the aggregate rewards their members receive. The result says little about the overall desirability or usefulness of the research, but much about priority of knowledge claims.

A Fable

We can explore the difference between foreseeable policies based on the reward system and institutional units, and those based on cognitive criteria with a thought experiment — or, more properly, a fairy tale. The tale rests on some substantial, but not totally unrealistic, assumptions about how science works; one of these is that the new discoveries of Big Physics are overdetermined, and another is that most top level scientists are fairly competent. Events in the fable are, of course, somewhat more neatly organized than they are in real life, but the principle still stands, I believe.

Imagine two laboratories — 'Paragon' cost $10M (Megabucks) per year to run, while 'Querulous' costs $9M per year. In a typical year Paragon produces ten major findings while Querulous produces none, merely rediscovering the findings of Paragon about six months later. Paragon publishes many important papers, its scientists win honours of various sorts, and it is highly esteemed in the scientific community. Querulous is generally agreed to be second rate; quantitative analyses of its output confirm this impression. It is, in IM's terms, a very ineffective laboratory when compared to Paragon. Taking the two laboratories as separate and individual entities we might conclude that to close down Querulous would save $9M per year for little loss in terms of discoveries in physics. But, if one thinks in terms of what will be discovered under various policies, instead of who has been most rewarded, and if one sets aside considerations of equity, it turns out that it could be more sensible to close Paragon!

This is how such an apparently counter-commonsensical conclusion can be reached. If we closed Paragon we could save not $9M but $10M per year. However, perhaps all the same discoveries would still be made, but it would simply all happen six months later. The same publication outlets would still need to be filled and, since the reward system would not change, the same prizes would be given away, but it would now be the scientist from Querulous who had the priority: they would be the recipients of their colleagues' esteem, and would score highly on IM's criteria. What one might call the 'Game Show Principle' — all the prizes must be given away, whatever the quality of the players — would ensure that there would be just as many prize winners as before.[6] (Actually, the metaphor is rather harsh, for Querulous's scientists are not dumb, just slightly slower than Paragon's. It is the reward system itself that makes tardiness look like incompetence.) Under these circumstances we would have the same science, a bit later, for an extra saving of $1M per year compared with the 'close Querulous' scheme. Under these assumptions it is the *absolute cost* of running the laboratories that matters for science policy, *not their relative cost-effectiveness*.

The story may not be realistic, but it makes the point of principle that treating science as an aggregate of individual laboratories, or an aggregate of individual scientists, does not reflect science as a knowledge-producing system. The problem is aggravated if scientific output is measured by indicators which reflect the reward system of science, with its atomistic and competitive ethos. In terms of purely cognitive goals — producing scientific findings — the institutional/atomistic approach can generate irrelevant impressions. Of course, the fable can give rise to no real policies. Equity is important — and, anyway, there is no political power that could close Paragon and leave Querulous alive. What is more, if Paragon were closed, Querulous would be costing at least $10M per

BEST COPY AVAILABLE

year to run before twelve months had elapsed. The point remains, however, that different ways of approaching the measurement of scientific quality can lead to widely different conclusions.

Conclusion

To end on a positive note, the last ten to fifteen years has seen substantial progress in studies that break up science by reference to cognitive, rather than institutional, criteria. The sociology and history of scientific knowledge have developed methodologies for analyzing passages of scientific activity, irrespective of their institutional locations. It is time that these studies started to be 'cashed in' for their policy implications. For example, we could begin to investigate the balance of the different 'phases of science' (normal and extraordinary) and the conditions of their survival;[7] the conditions for the maintenance and transmission of bodies of scientific skill; the way that the allocation of research funds affects the practice and findings of scientific research; and the way in which the scientific community constrains, or encourages, the production of novelty. While quantification of these exercises will be difficult, we should not allow our attention to be distracted from the serious problems of number fetishism — as John Irvine must know.[8] Were these important goals to be pursued with IM's energy, we might develop a science policy that would do more than offer short-term solutions to perceived cost-ineffectiveness.

• *NOTES*

This is a modified version of part of a paper first presented at the Science Studies Committee meeting at Imperial College, London, on 25 June 1983. Another part of that paper has been published in *EASST Newsletter*, No. 2 (November 1983), Vol. 4, 5–8. I am grateful to Jay Gershuny and John Cullis for very helpful comments on an earlier draft of this Response.

1. As IM point out, there may be 'non-cognitive' aims for science policy, too. For example, economic and technological spin-offs from scientific activity might be seen as more important than the science itself; scientific training is vital, and prestige and political considerations are not to be ignored. However, a policy which was directed toward maximizing some or all of these goals would not use IM's methods, either.

2. J. Irvine and B. Martin, 'Basic Research in the East and West: A Comparison of High-Energy Physics Accelerators', *Social Studies of Science*, Vol. 15 (1985), 293–341.

3. See, for example, J. Irvine and B. Martin, 'Assessing Basic Research: The Case of the Isaac Newton Telescope', *Social Studies of Science*, Vol. 13 (1983), 49–86; Martin and Irvine, 'An Evaluation of the Research Performance of Electron High-Energy Physics Accelerators', *Minerva*, Vol. 14 (1981), 408–32; Martin and Irvine, 'Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, Vol. 12 (1983), 61–90.

4. A general criticism of 'methodological triangulation' is that it is not clear what one does when the measures do not converge. I agree with Krige and Pestre that non-convergence is the more interesting and demanding case. As will be seen below, I do not agree that IM's measures converge on scientific quality in any straightforward way, since priority determines rewards to too great an extent. See J. Krige and D. Pestre, 'A Critique of Irvine and Martin's Methodology for Evaluating Big Science', *Social Studies of Science*, Vol. 15 (1985), 525–39.

5. P. Medawar, *The Art of the Soluble* (London: Penguin, 1969), 96.

6. Some economists would say that scientific rewards are a 'positional good'.

7. See H. M. Collins, *Changing Order: Replication and Induction in Scientific Practice* (London and Beverly Hills, Calif.: Sage, 1985).

8. See J. Irvine, I. Miles and J. Evans, *Demystifying Social Statistics* (London: Pluto Press, 1979).

Author's address: Science Studies Centre, School of
Humanities and Social Sciences, University of Bath, Claverton
Down, Bath BA2 7AY, UK.

Responses and Replies (continued)

Evaluating the Evaluators: A Reply to Our Critics

Ben R. Martin and John Irvine[1]

We should begin by welcoming the opportunity to have our work critically reviewed in this journal. Given the wide range of differences we have with our critics, it is perhaps first worth identifying the issues on which we are in agreement. Four stand out in particular:

Social Studies of Science (SAGE, London, Beverly Hills and New Delhi), Vol. 15 (1985), 558–75

(1) research addressing science policy issues is urgently needed;

(2) advances in the area of science studies over the last two decades mean that it is now possible to tackle substantive science policy issues in a more systematic manner;

(3) bibliometric studies have begun to focus on 'issues of importance' and 'have made a marked impact',[2] even succeeding in certain cases in providing 'a wealth of statistical data to be taken into account by future policy makers';[3]

(4) our approach is certainly capable of being improved upon.

In view of the unanimity on these points, we are disappointed at how little our critics have to offer by way of concrete suggestions as to how one might improve the techniques we have been attempting to develop to provide systematic and reliable information for science policymaking. Let us consider in turn the points made in the four critiques and the improvements (if any) in evaluation methods and approaches that they suggest.

Policy Research or Fairy Tales?

Harry Collins begins by congratulating us for provoking 'some serious questioning of British science policy'. 'Nevertheless', he continues, 'their research is not without flaws, nor is it obviously the best way for science policy to make use of recent work in science studies'.[4] What are these 'flaws'?

The first is that we have taken as our unit of analysis research institutions rather than units reflecting cognitive boundaries.[5] According to Collins, policy research should 'start by disaggregating science according to cognitive rather than institutional boundaries'.[6] This is a fundamental difference, and one on which we disagree strongly with Collins. In basic science, at least, policymakers often take decisions on whether to fund research groups, departments, laboratories, and centres, as opposed to entities defined purely by cognitive boundaries. While the latter may be the most suitable unit of analysis for the study of certain questions in the sociology of science, for science policy research the former is important and clearly cannot be ignored. Examination of the expenditure of most national research-funding agencies shows that research groups, laboratories, and so on — that is, institutionally defined entities — account for a large proportion of their total budgets (some grants are, of course, made to individuals — most importantly to open up new areas — but these account for only a fraction of the funds). There are, of course, other cases, generally involving more applied research, where such organizations may have to decide whether or not to support a field (and in our more recent work we have tried to address this type of question[7]). However, even then, once an agency has decided to support a field for strategic or other reasons, it still has to choose which research groups or institutions to support, so information relating to their past and likely future performance is just as vital. In short, Collins is wrong to assume that what is most interesting from the point of view of a particular sociology of science standpoint coincides exactly with the main interests of policy-makers. To paraphrase Collins, 'Science policy should be about maintaining efficient institutions *as well as* maintaining the sort of cognitive community that will produce the desired scientific products'. (As an aside, we would also point out that much interesting sociology of science would be lost if Collins' demarcation were to be universally adopted.)

The second criticism is that the indicators we have employed reflect the atomistic reward system of science, with all the prizes going to those who are first. In fact, only one of our indicators (highly-cited papers) relates directly to 'discoveries' or 'coming first' — the others relate much more to taking 'part in a scientific endeavour'.[8] For example, our study of world high-energy physics showed that between 1961 and 1982 there were very few occasions on which CERN was 'first' (nearly all the crucial discoveries and advances over this period were made in the United States).[9] However, the other indicators suggest that CERN accelerators were perhaps the most successful in terms of experiments yielding more precise measurements and better statistics, a finding in line with the results obtained from 180 interviews.[10]

Collins then attempts to demonstrate by means of 'a fairy tale' 'the difference between foreseeable policies based on the reward system and institutional units, and those based on cognitive criteria'.[11] The tale, we are informed, 'rests on some substantial, but not totally unrealistic, assumptions about how science works'. One assumption (apparently unrecognized by Collins) is that the international dimension of science can be completely ignored. (This assumption may actually hold in Collins's own research area, but that is something of an exception.) Besides Paragon and Querulous in country (or continent) A, there would be other Paragons and Querulouses around the world. Hence, if a British Research

BEST COPY AVAILABLE

Council, for example, were to decide to close the world leader in a particular field, the probable result would not be that the same discoveries would be made six months later by a less costly UK group, but by Texas Megabuck Lab or Nippon Lab. Furthermore, Collins's fairy tale, even if it had been based on realistic assumptions, entirely misses the point. In most cases, those *outside* the field concerned will not know that Paragon has been making ten major advances a year while Querulous has been making none. Scientists within the field will undoubtedly have subjective impressions of the difference between the two, but, for reasons related partly to the existence of vested interest groups, they may not reveal the true extent of the difference to others. Nor are they likely to have as wide-ranging an understanding of the factors structuring research performance as that, for example, yielded in our study of high-energy physics accelerators. We have seen our task as one of providing systematic information on research performance and the factors structuring it, in a form accessible not just to researchers in the field concerned but also to policymakers, scientists in other fields, politicians and the public, thereby making possible greater transparency in the scientific decision making process.

Having argued that our work is not 'the best way for science policy to make use of recent work in science studies'[12] (and employed his fairy tale to prove that 'different ways of approaching the measurement of scientific quality can lead to widely different conclusions'[13]), what does Collins suggest might constitute 'the best way' forward? As his first example, he proposes that 'we could begin to investigate the balance of the different "phases of science" (normal and extraordinary) and the conditions of their survival'.[14] But does this (or indeed any of the other examples Collins cites) actually correspond to the primary concerns of policymakers? Over the last seven years, we have devoted considerable effort to trying to establish just what are their main interests and problems. Almost all the many policymakers in the UK and overseas who have discussed this with us have stressed the need for better information on the performance of research groups, departments, facilities and laboratories, and on where their country stands in international terms in particular fields.[15] As far as we can recollect, not one has mentioned any of the examples listed by Collins.[16]

The Case of the Misread Papers

It is difficult to know how to take the criticisms by Robert Bud when there appears to have been such a fundamental misreading of our work as to lead him to categorize high-energy physics as 'strategic research'. In our book, *Foresight in Science*, we explain why the conventional classification of R&D (enshrined in the OECD 'Frascati Manual') into 'basic research', 'applied research' and 'experimental development' is increasingly inadequate. In particular, within basic research one can now identify two very different types of research activity, namely 'curiosity-oriented research' and 'strategic research' — the latter being work where no specific end-product or process can yet be identified (so it is not 'applied' research) but where the research is expected to produce a broad base of knowledge likely to form the background to the solution of practical problems. The book then concentrates on methods for looking at the longer-term future for strategic research. It does not, as Bud claims, dismiss curiosity-oriented research as an empirically residual class, but rather argues that its greater intrinsic unpredictability makes attempts at longer-term forecasts less reliable and therefore not so worthwhile. As with any categorization, 'curiosity-oriented research' and 'strategic research' are to some extent ideal types, but it is not difficult to think of examples of research activities which are predominantly one or the other. Most basic research biotechnology, for example, is primarily 'strategic', while at the other end of the spectrum we would place almost all astronomy and high-energy physics. We can only assume that Bud's view of high-energy physics as strategic research is linked to his belief (stated in two places) that experimental high-energy physics is carried out on 'reactors'[17] and therefore is presumably in some way part of nuclear energy research. In fact, the two fields parted company in the 1950s, and the links between them (both cognitive and social) are now almost non-existent (or at least no greater than with other fields of physical science). This is the reason why in our study we concentrated on evaluating the performance of accelerators in scientific terms. We would, however, welcome positive suggestions on how to extend the evaluation to cover 'extrinsic' goals.[18]

Putting aside the doubts engendered by such a mistake, what can we make of Bud's criticisms? We are told that we have devoted insufficient attention to examining the goals of scientific research (although Bud does not actually tell us what he thinks these are), and that we have used indicators 'which do not indicate levels of achievement of those hypothetical [sic] goals'.[19] Right from the start of our research programme, we have explicitly recognized that one cannot use the same indicators to evaluate different types of research. If Bud were to read, for example, the report on a study of mechanical, electrical and electronics engineering that we carried out for a 1981 Norwegian Royal Commission, he would see that a very different approach and set of indicators was adopted to evaluate these more applied areas, whose objectives clearly differ from those, say, in high-energy physics. (We laid great stress, for example, on 'customer review' — that is, obtaining the views of industrial firms on the research activities being assessed — and produced a large volume of quantitative data relating to this.)[20]

A second criticism relates to our comparison of the research performance of accelerators in the Eastern bloc, the United States and Western Europe, with Bud pointing out that the accelerators 'are embedded in three such different cultural systems'.[21] Whereas one response to these cultural differences may be to throw up one's arms in horror and imply that no valid comparisons can apparently ever be drawn, ours is to attempt to identify variations in scientific output and impact and then examine the extent to which these can be related to the cultural differences. (Eastern and Western high-energy physicists are not in fact the two completely distinct and non-interacting communities that Bud implies. They attend many of the same conferences, use each other's accelerators to a limited extent, and publish in essentially the same body of international journals, with East Europeans having made increasing use of West European journals over the last twenty years.)

Next, Bud chides us that 'the interpretation of the quantitative data is perfunctory. Conclusions are drawn only about overall magnitudes'.[22] As with any piece of research (but particularly one in a new area), the analysis could have been taken further. We have not attempted to solve all the problems in making East-West comparisons of scientific performance at once and have also been careful to avoid drawing over-ambitious conclusions (that is to say, the caveat has not 'disappeared' as Bud claims) which cannot be adequately supported by data with all the limitations we describe. If Bud has specific suggestions to make in relation to interpreting the empirical data further, we would appreciate the opportunity to consider them.

Overall, Bud, like Collins, seems to exhibit a myopic tendency to see everything from the perspective of science studies rather than science policy. His 'problem with all the indicators is the same. What are their implications for our understanding of the relationship between the scientific communities?'[23] We would merely note that this is not a problem that science policymakers have identified to us as one of their principal concerns.

The CERN Critique

The stated aim of the critique by John Krige and Dominique Pestre is simple and admirable — it is to defend 'intellectual rigour'.[24] It is therefore alarming to find them beginning the presentation of their case by quoting what someone stated at a seminar had been said to him on some other unspecified occasion by 'a Chinese colleague' who was in turn 'quoting' two other people. (Did Marx and Engels actually both say anything as simplistic as that — we would be interested in seeing the relevant references!) Yet, later in their Response, this unsubstantiated anecdote has been elevated to the status of evidence which 'makes the point'. Is this more 'rigorous' than our structured interviews with over 180 high-energy physicists, each of whom then completed a detailed attitude survey?

After presenting a clear summary of our work on big science, Krige and Pestre commence their critique by arguing that such non-scientific benefits as 'manpower' training, technological spin-off and national prestige are not 'secondary reasons' for supporting high-energy physics. This would seem to suggest that they see such benefits as 'primary' — that is, of equal (or greater?) importance than the scientific question of how much the research will increase our knowledge of the material world. (Krige and Pestre do actually contradict this two pages later: 'We are not claiming that "non-scientific" considerations, rather than the production of scientific knowledge, are "primary" or "more important" in assessing Big Science.'[25] But if they are neither 'primary' nor 'secondary', what exactly are they? A little more rigour is required here.) This raises the interesting (and unanswered) question, 'How poor would the scientific performance of CERN have to become before scientific criteria finally over-rode non-scientific ones?' Or are we to agree with Moed and van Raan that centres like CERN 'are simply too rare, too prestigious' ever to close down?[26]

In concluding our analysis of CERN's past performance, we discussed the political, educational and technological benefits it has yielded.[27] We are, however, less convinced than our CERN colleagues by the findings of Schmied on technological spin-off from CERN.[28] While it may be comforting to CERN to have a study claiming to find that the industrial contracts placed by CERN have resulted in increased 'economic utility' several times greater than the original value of those contracts, this should not blind us to *its* methodological inadequacy. First of all, the concept of 'economic utility' employed in the study would be unrecognizable to most economists. Secondly, the methodological approach of asking firms to identify subsequent contracts that have drawn upon work originally carried out for CERN and then counting their total value as an economic benefit from CERN is extremely dubious. Thirdly, there is the question of how much reliability one can place on the answers of firms dependent to a greater or lesser extent on CERN for future business when questioned in a study financed by CERN. Lastly, the opportunity costs are completely ignored — 'is the level of technological spin-off higher than it would have been if the resources spent on CERN had instead been used to support some other type of research, such as exploration of the ocean bed, for example?'[29] Given these various doubts, we would place rather less confidence in the results of this study than do Krige and Pestre.

While we recognize the importance of the various educational, technological and political benefits associated with CERN, and would have devoted more effort to evaluating CERN in terms of them if time and resources had permitted (we certainly do not dismiss them as 'trivial' as Bud claims[30]), we would still argue that CERN in particular, and Big Sciences like astronomy and high-energy physics in general, are supported more for the contributions they are expected to make to scientific knowledge.[31] In our papers on CERN,[32] we quote various explicit statements from CERN to this effect.[33] Furthermore, few of the scientists involved attempt to defend the funding for their work by reference to 'extrinsic' benefits: our attitude survey revealed that twice as many felt that expenditure on the field could be justified only in scientific terms as believed it could in terms of the spin-off it generates. Krige and Pestre's observation that Big Sciences like high-energy physics 'have a kind of role analogous to that of military research in the economical field' is perhaps more revealing than they intend, given the extensive literature demonstrating how military R&D, while it does yield occasional important spin-offs, at the same time locks up scarce R&D resources that might more profitably be employed in other forms of research.[34]

The next point made by Krige and Pestre is that we 'rely essentially on the *average feelings* of the scientific community, leading us to doubt whether they can produce assessments substantially different from those obtained through a more classical panel system'.[35] Here, it is important to distinguish between conventional peer-review (involving a small number of referees or 'experts' on a panel) and our extensive peer-evaluations drawing in very large numbers of researchers across different countries and based on structured confidential interviews and attitude surveys. Because the latter approach yields relatively consistent results, it does not logically follow that conventional peer-review is 'rather reliable'.[36] Indeed, we have encountered several instances where the existence of an 'oligopolistic' situation in a research field has led to the two approaches yielding very different results. One of the most prominent (and costly) involved the ISABELLE project in the United States,[37] where conventional peer-review continued to suggest that this was the top priority for US high-energy physics long after most researchers had privately recognized that this was no longer true (and admitted as much to us in interviews). By the time the project was finally aborted, some $200m had been spent.

Another important problem with conventional peer-review is the inherent lack of accountability to those outside the specialty concerned. While it may be easy for a panel of high-energy physicists, for example, to conclude that existing accelerators are successful and that a proposed new facility is an absolute priority, such statements can equally easily be dismissed by outsiders on the 'Rice-Davies' grounds that 'They would say that, wouldn't they?' Hence the need to complement traditional peer-review with bibliometric data on research performance, or external assessments of the likely future performance of new facilities. If presented in a form accessible to those outside the field, such information can play a major role in providing evidence to other scientists competing for scarce funds (as well as to policymakers and the general public) that decisions are well founded. In this way, one can begin to achieve more open and transparent decision-making in basic science.

Krige and Pestre, in fact, provide a good example where such accountability may have been somewhat lacking. At the time it was built, the CERN SPS accelerator was criticized by some, particularly in the United States, for being over-engineered and 'gold-plated' (it cost almost twice as much as the similar-energy Fermilab accelerator). Yet, as Krige and Pestre point out, it was this which permitted the SPS to be converted into a proton-antiproton collider more quickly and at lower cost than the Fermilab machine, and hence to discover the Intermediate Vector Bosons. However, this raises several awkward questions: how many other instances of 'gold-plating' have there been which have not paid off in this way? At what cost? And were CERN member states aware, when they agreed to the SPS project, that they were funding a comparatively expensive machine which might or might not turn out to have advantages over a cut-price version? As defenders of intellectual rigour, our critics must recognize that one example where 'gold-plating' did pay off by no means proves that this policy always constitutes the best use of limited resources.

As for the 'rhetorical rules' identified by Krige and Pestre, we take these as not wholly uncomplimentary. We would, for example, prefer to be accused of appealing to common sense than the alternative, whatever that might be. However, on a point of information in relation to the first rule, we do not '*envisage* all possible objections first',[38] in the sense of dreaming them up ourselves beforehand. At the end of the CERN study, we spent a year circulating drafts of the papers for comment to around 100 high-energy physicists, other scientists and officials in funding agencies, then revising and recirculating them. We also debated the results with the scientific community in various seminars. This was the way in which possible objections were identified, and then taken into account before submitting the papers for publication. It is our view that work on science policy must be capable of being defended in front of a scientific audience. Given the emphasis that we have placed on such validation of our results by the scientific community, we would dispute the claim by Moed and van Raan that there is no test of the 'validity' of our method.[39]

The Leiden Alternative?

This brings us to the criticisms of Henk Moed and Anthony van Raan. The first relates to the various indicators that we have employed and the extent to which they converge. The indicators, although related to some extent, do nevertheless reflect slightly different facets of research performance. Publication totals give an indication of the overall scientific production of a research group, while numbers of papers per researcher or per dollar reveal something about its productivity. The average number of citations per paper is a measure of the impact those publications have on the scientific community, while peer-rankings (where scientists rank the contributions of different institutions) provide evidence on the perceived significance of the results. Lastly, data on the distribution of highly-cited papers reveal which groups have been responsible for the few key 'discoveries' or advances in a specialty, while aggregate citation statistics reflect the overall impact of a large number of incremental additions to knowledge. (Contrary to what some critics suggest, we do not claim that any of the indicators measure the 'quality' of research.[40]) As we have spelt out in detail,[41] all these indicators are 'partial' in nature. However, in the real world of science policy, one has to accept that there are no perfect measures of scientific performance and use whatever indicators are available, recognizing their limitations and working with them as best as possible.

The starting point when we began work on research evaluation was the hypothesis that, if the various indicators were applied to matched research groups, we would expect them to yield broadly convergent results. This they did in the case of electron high-energy physics, radio astronomy and optical astronomy. In the latter, for example, we obtained the results summarized in Table 1. We are surprised by Moed and van Raan's claim[42] that the indicators do not converge in this instance. As we stated when reporting the very first research evaluation results, the fact that the indicators converge in a given case does not 'prove' that the results are 100 percent certain — the indicators may all be 'wrong' together.[44] However, if a research facility like the Lick 3-metre telescope produces a comparatively large publication output at fairly low cost, if those papers are relatively highly cited, if it yields many of the highly-cited papers in the field, and if large numbers of astronomers rate it highly in the course of structured interviews, we would place *more* credibility on the resulting conclusion that this was a successful facility than if the same finding was arrived at by a panel of three or four 'experts' without access to the sort of systematic information that we have collected. Similarly, if a telescope like the INT produces relatively few papers at high cost which receive comparatively few citations in total (even though each paper on average has a reasonable citation-

BEST COPY AVAILABLE

TABLE 1
Output Indicators for Optical Telescopes — A Summary[43]

	Lick 3-metre	KPNO 2.1-metre	CTIO 1.5-metre	INT 2.5-metre
Average no. of papers p.a., 1969–78	42	43	35	7
Cost per paper in 1978	£13k	£7k	£6k	£63k
Citations to work of past 4 years in 1978	920	710	580	140
Av. citations per paper in 1978	4.2	3.3	3.3	3.6
No. of papers cited 12 or more times in a year, 1969–78	41	31	21	4

per-paper figure), and if it yields rather few highly-cited papers and is ranked towards the bottom of a list of 12 telescopes by 50 astronomers, we would be reasonably confident that its scientific performance had not been particularly good in world terms. Even so, we have always stressed that such findings need to be interpreted with care, and that, rather than being used to replace the peer-review process, they should be fed into it to enhance its effectiveness and help keep decision-making 'honest'.

As for the concept of 'convergence', like any notion when first formulated it may initially have been somewhat 'poorly developed'. However, as more empirical studies have been completed, so the conditions under which the indicators might be expected to converge have become clearer. First, the less satisfactorily research groups are 'matched', the less convergence is to be expected. The convergence for the four radio astronomy centres was reasonably good — certainly much better than when we compared CERN (which then had three major machines) with other high-energy physics laboratories (mostly operating just one accelerator each). Secondly, in a period of revolutionary change within a field, the facility

BEST COPY AVAILABLE

producing the most highly-cited papers and 'discoveries' (as SLAC did in experimental high-energy physics during the mid-1970s) is likely to be judged the most successful even if it is not the world leader in terms of numbers of publications and citations. Thirdly, convergence may be expected if the indicators are applied to research groups working in an *internationally homogeneous* field, but not if the field consists instead of several distinct, non-interacting communities (which do not attend the same conferences, publish in the same journals, cite each other's work, and so on).[45] For situations between these extremes, where there is some interaction but not complete international homogeneity (as we encountered when comparing high-energy physics in the East and West[46]), one may, for example, first have to adjust the indicators for differences in publication and referencing practices. This is, however, an area where more work is needed.

Another disagreement we have with our Dutch critics concerns the respective merits of computerized and manual-scanning approaches to research evaluation. We have no objection to computing *per se*, and indeed use outside data-bases where there are significant advantages — for example, the *NSF Science Literature Indicators Data-Base*, which we have used to draw conclusions about the overall scientific performance of countries across research fields.[47] However, there are certain tasks in research evaluation where a manual scanning approach is the only option:

(1) In some cases, computerized scanning is too expensive because of the high access costs to data-bases (research evaluations should not cost more than a fraction of the research being evaluated).[48]
(2) There can be difficulties with using existing computerized data-bases in defining the boundaries of fields.[49] In the case of the CERN study, for example, it was necessary to construct our own data-base on experimental high-energy physics. Here, there was no option to a manual-scanning approach because of the need to read large numbers of physics papers in order to establish which related to experimental high-energy physics and which did not.
(3) Without reading the papers, it is generally impossible to establish which research facility has been used to produce the experimental results (this is necessary for carrying out analyses of institutional and/or national performance).
(4) Similarly, reading is essential to distinguish between experimental and theoretical papers in a given field, something which is vital from a policy point of view for areas where the costs of experimental and theoretical work are very different.

It is therefore misleading to pretend that the choice between manual and computerized scanning 'all depends on the search-algorithm'.[50]

For the tasks mentioned above, the best search-software is still to be found in the human brain. Furthermore, the worry about 'the scanner getting tired (or even crazy)'[51] needs to be counterbalanced by the benefits associated with actually reading the papers and becoming immersed in their content.

Since Moed and van Raan hold up their own work as an exemplar of bibliometric assessment, it should be pointed out that their approach is in our view somewhat flawed in that performance indicators are applied to different departments within a single university — that is, they are used to draw comparisons between fields characterized by very different publication and citation practices.[52] There is no attempt to compare 'like' with 'like'. While it may be relatively simple for those with sufficient funds to obtain bibliometric data for such groups (by ordering computer print-outs from commercial data-banks), we harbour grave misgivings about applying indicators in such a fashion, and would be interested to know what the researchers thus assessed think of the validity of the results.

Moed and van Raan's final criticism is that 'performance analyses of large facilities such as accelerators have a limited relevance for research policy'.[53] In their view, 'as soon as, at least in the West, a specific facility — like an accelerator — loses its position at "the front", scientists will move to other facilities. . . . It would . . . be [more] relevant to policy to follow specific (groups of) scientists . . . These groups . . . are, in our opinion, the most interesting (and therefore most policy-relevant) "level" to evaluate.'[54] This criticism is misplaced for two reasons. First, it assumes a perfect labour market within science completely at odds with the actual situation (even in the West).[55] Secondly, central facilities account for a large proportion of the expenditure by national research-funding agencies,[56] and for Big Sciences like high-energy physics the main policy decisions focus on whether to fund a centre or a new accelerator or detector, not on which university user groups to support.

Concluding Note

Let us conclude as we began by returning to an area where we are in agreement with our critics. Collins ends his paper by observing that, given the substantial progress in science studies over the last ten to fifteen years, it is time for the results to be 'cashed in' for their policy implications. We agree entirely, although it is important that policymakers are not 'short changed' in the process. We would, however, offer two further notes of caution. First, as with any 'applied' research, prior 'market research' is absolutely essential — in this case, to establish what are the needs of science policymakers. It cannot be assumed that what is most interesting from a sociological point of view is necessarily most relevant from a policy perspective. Secondly, sociologists and others in the science studies community will have to do rather more than weave fairy tales if they are to convince policymakers and scientists that their research has some validity and the results it yields some utility. In particular, their work must be capable of being defended in front of a scientific audience, where it will be judged in terms of criteria somewhat different from those normally employed in science studies.

Finally, on a more personal note, we warmly welcome the fact that Harry Collins, in discussing what constitutes 'the *best* way for science policy to make use of recent work in science studies'[57] seems at long last to be renouncing, implicitly at least, ultra-relativism. We applaud this courageous step 'forward', since science policy could potentially derive great benefit from the qualitative sociology of science if the two sides were to work more closely together than hitherto.

• *NOTES*

The authors wish to thank the Leverhulme Trust, which currently funds a three-year programme on research evaluation at the Science Policy Research Unit, for the financial support required in preparing this paper. They are also grateful to Andrew Barry, Diana Hicks, Ian Miles, Geoff Oldham, Keith Pavitt and John Ziman for helpful comments on an earlier draft of this paper.

1. No order of seniority implied (rotating first authorship).

2. J. Krige and D. Pestre, 'A Critique of Irvine and Martin's Methodology for Evaluating Big Science', *Social Studies of Science*, Vol. 15 (1985), 525–39, quote at 525.

3. R. Bud, 'The Case of the Disappearing Caveat: A Critique of Irvine and Martin's Methodology', *Social Studies of Science*, Vol. 15 (1985), 548–53, quote at 548.

4. H. M. Collins, 'The Possibilities of Science Policy', *Social Studies of Science*, Vol. 15 (1985), 554–58, quote at 554.

5. We are puzzled that, where we have focused on an entire field or discipline, this should still be classified by Collins as a unit with a 'non-cognitive' boundary.

6. Collins, op. cit. note 4, 554–55.

7. See J. Irvine and B. R. Martin, *Foresight in Science; Picking the Winners* (London: Frances Pinter, 1984), and Martin, Irvine and Turner, 'The Writing on the Wall for British Science', *New Scientist* (8 October 1984), 25–29.

8. Collins, op. cit. note 4, 555.

9. See J. Irvine and B. R. Martin, 'CERN: Past Performance and Future Prospects — II. The Scientific Performance of the CERN Accelerators', *Research Policy*, Vol. 13 (1984), 247–84.

10. Ibid., 281.

11. Collins, op. cit. note 4, 555.

12. Ibid., 554.

13. Ibid., 557.

14. Ibid.

15. See also *An Agenda for a Study of Government Science Policy*, a report prepared by the Task Force on Science Policy for the Committee on Science and Technology of the US House of Representatives, Ninety-Eighth Congress, Second Session (Washington, DC: US GPO, 1984). This lists some 140 science-policy related questions.

16. The points about 'non-convergence' raised by Collins, op. cit. note 4, 558, fn. 4, are discussed in the section below dealing with the criticisms of Moed and van Raan.

17. Bud, op. cit. note 3, 550 and 552.

18. For a response to Bud's criticism that we dismiss external (or non-scientific) benefits from high-energy physics as 'trivial', see the following section dealing with the comments of Krige and Pestre.

19. Bud, op. cit. note 3, 548.

20. See J. Irvine, B. R. Martin and M. Schwarz, with K. Pavitt and R. Rothwell, *Government Support for Industrial Research in Norway: A SPRU Report* (Oslo: Universitetsforlaget, Norwegian Official Publication NOU 1981: 30B, 1981).

21. Bud, op. cit. note 3, 548.

22. Ibid., 552.

23. Ibid., 552–53.

24. Krige and Pestre, op. cit. note 2, 526.

25. Ibid., 529.

26. J. F. Moed and A. F. J. Raan, 'Critical Remarks on Irvine and Martin's Methodology for Evaluating Scientific Performance', *Social Studies of Science*, Vol. 15 (1985), 539–47, quote at 544–45.

27. See Irvine and Martin, op. cit. note 9, 281.

28. H. Schmied, 'A Study of Economic Utility Resulting from CERN Contracts', *IEEE Transactions in Engineering Management*, Vol. EM-24 (1977), 125. This Study has recently been updated by Schmied and various colleagues at CERN — see M. Bianchi-Streit, N. Blackburne, R. Budde, H. Reitz, B. Sagnell, H. Schmied and B. Schorr, *Economic Utility Resulting from CERN Contracts (Second Study)* (Geneva: European Organization for Nuclear Research, CERN 84–14, 1984).

29. Irvine and Martin, op. cit. note 9, 281.

30. Bud., op. cit. note 3, 549.

31. See the discussion in J. Irvine and B. R. Martin, 'The Economic Effects of Big Science: The Case of Radio Astronomy', *Proceedings of the International Colloquium on Economic Effects of Space and other Advanced Technologies, Strasbourg, 28–30 April 1980* (Paris: European Space Agency, ESA SP-151, 1980).

32. B. R. Martin and J. Irvine, 'CERN: Past Performance and Future Prospects — I. CERN's Position in World High-Energy Physics', *Research Policy*, Vol. 13 (1984), 183–210; and Irvine and Martin, op. cit. note 9.

33. Even Schmied and his colleagues agree that 'The primary function of CERN is to carry out "very basic research" in particle physics, and the direct product of this research work is scientific knowledge or "culture"': Bianchi-Streit et al., op. cit. note 28, 1.

34. See, for example, M. Kaldor, 'Technical Change in the Defence Industry', in K. Pavitt (ed.), *Technical Innovation and British Economic Performance* (London: Macmillan, 1980), 100–25; K. Dickson, 'The Influence of Ministry of Defence Spending on Semiconductor Research and Development in the United Kingdom', *Research Policy*, Vol. 12 (1983), 113–20.

35. Krige and Pestre, op. cit. note 2, 530 (emphasis in original).

36. Ibid., 532.

37. For details, see B. R. Martin and J. Irvine, 'CERN: Past Performance and Future Prospects — III. CERN and the Future of World High-Energy Physics', *Research Policy*, Vol. 13 (1984), 311–42.

38. Krige and Pestre, op. cit. note 2, 536 (emphasis added).

39. Moed and van Raan, op. cit. note 26, 542.

40. See the distinction between 'quality', 'importance' and 'impact' made in B. R. Martin and J. Irvine, 'Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, Vol. 12 (1983), 61–90.

41. Ibid., 65–74.

42. Moed and van Raan, op. cit. note 26, 544, and again at 541.

43. Taken from Tables 3 and 4 in J. Irvine and B. R. Martin, 'Assessing Basic Research: The Case of the Isaac Newton Telescope', *Social Studies of Science*, Vol. 13 (1983), 49–86.

44. Martin and Irvine, op. cit. note 37, 87, fn 55.

45. This may explain the claim by Moed and van Raan (op. cit. note 26, 546) that this [illegible]ns-per-paper indicator does not allow for differences in scale of research activity. While this [illegible] may be true for fields characterized by localized, non-interacting communities with differing propensities to engage in self-citation and in-house citation, for a reasonably internationally homogeneous field like high-energy physics, the citations-per-paper figure for a particular facility is much less dependent on its publication rate.

46. Moed and van Raan are wrong in concluding that we have assumed that Eastern-bloc papers refer *exclusively* to other Eastern-bloc papers (ibid., 545). What we have allowed for is the fact that Eastern papers lose *far more* citations from unscanned journals that do Western ones, and we did this by taking a sample of references in Eastern papers and examining the proportions given to Eastern and Western papers.

47. See Martin, Irvine and Turner, op. cit. note 7.

48. This was a major point of discussion in the UK Advisory Board for the Research Council's Science Policy Study 1983/84, part of which is described in B. R. Martin, J. Irvine and D. Crouch, *Science Indicators for Research Policy: A Bibliometric Analysis of Ocean Currents and Protein Crystallography* (Brighton: University of Sussex, Science Policy Research Unit Occasional Paper No. 23, 1985).

49. For details, see ibid.

50. Moed and van Raan, op. cit. note 26, 545.

51. Ibid.

52. See H. F. Moed, W. J. M. Burger, J. G. Frankfort and A. F. J. van Raan, *On the Assessment of Research Performance: The Use of Bibliometric Indicators* (University of Leiden: Research Policy Unit, 1983).

53. Moed and van Raan, op. cit. note 26, 546–47.

54. Ibid., 544.

55. CERN member states like Britain expect the bulk of their high-energy physicists to use CERN facilities (only very limited resources are available to fund experiments at US laboratories, for example).

56. See J. Irvine and B. R. Martin, 'What Direction for Basic Scientific Research?', in M. Gibbons, P. Gummett and B. M. Udgaonkar (eds), *Science and Technology Policy in the 1980s and Beyond* (Harlow, Essex: Longman, 1984), 67–98.

57. Collins, op. cit. note 4, 554 (emphasis added).

Authors' address: Science Policy Research Unit, University of Sussex, Mantell Building, Falmer, Brighton, East Sussex BN1 9RF, UK.

BEST COPY AVAILABLE